

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 12:28:25 ; Search time 167 Seconds
(without alignments)
2004.080 Million cell updates/sec

Title: US-10-600-070B-2

Perfect score: 4063

Sequence: 1 MEALSHVIGLSPPQLCRLP.....YEVFWSKSGWKITEGSLVAS 801

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap:*

2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap:*

3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pap:*

4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pap:*

5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pap:*

6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|----------------------|
| 1 | 4063 | 100.0 | 801 | 4 | US-10-600-070-2 |
| 2 | 4051 | 99.7 | 801 | 4 | US-10-600-070-127 |
| 3 | 4051 | 99.7 | 801 | 4 | US-10-600-070-129 |
| 4 | 4051 | 99.7 | 801 | 4 | US-10-600-070-131 |
| 5 | 4051 | 99.7 | 801 | 5 | US-10-739-930-5771 |
| 6 | 2926 | 72.0 | 578 | 4 | US-10-600-070-123 |
| 7 | 1775.5 | 43.7 | 760 | 4 | US-10-600-070-125 |
| 8 | 1646 | 40.5 | 324 | 4 | US-10-600-070-206 |
| 9 | 1332 | 32.8 | 525 | 4 | US-10-600-070-122 |
| 10 | 1134 | 27.9 | 344 | 4 | US-10-424-599-271849 |
| 11 | 973 | 23.9 | 364 | 4 | US-10-425-115-266516 |
| 12 | 972 | 23.9 | 480 | 4 | US-10-437-963-172416 |
| 13 | 856 | 21.1 | 416 | 4 | US-10-425-115-242115 |
| 14 | 524.5 | 12.9 | 768 | 4 | US-10-600-070-167 |
| 15 | 499 | 12.3 | 204 | 4 | US-10-424-599-177901 |
| 16 | 498.5 | 12.3 | 789 | 4 | US-10-600-070-194 |
| 17 | 497 | 12.2 | 798 | 4 | US-10-600-070-164 |
| 18 | 497 | 12.2 | 798 | 4 | US-10-600-070-165 |
| 19 | 482 | 11.9 | 709 | 4 | US-10-437-963-172415 |
| 20 | 474.5 | 11.7 | 1157 | 4 | US-10-437-963-155799 |
| 21 | 442 | 10.9 | 631 | 4 | US-10-600-070-5 |
| 22 | 442 | 10.9 | 631 | 4 | US-10-600-070-162 |
| 23 | 435.5 | 10.7 | 191 | 4 | US-10-767-701-35633 |
| 24 | 434.5 | 10.7 | 624 | 4 | US-10-600-070-116 |
| 25 | 431 | 10.6 | 819 | 4 | US-10-600-070-171 |
| 26 | 431 | 10.6 | 819 | 4 | US-10-600-070-173 |
| 27 | 429 | 10.6 | 714 | 4 | US-10-600-070-169 |

| | | | | | | |
|----|-------|------|-----|---|----------------------|--------------------|
| 28 | 429 | 10.6 | 714 | 4 | US-10-600-070-170 | Sequence 170, App |
| 29 | 425.5 | 10.5 | 652 | 4 | US-10-600-070-115 | Sequence 115, App |
| 30 | 423 | 10.4 | 673 | 4 | US-10-600-070-192 | Sequence 192, App |
| 31 | 413 | 10.2 | 836 | 4 | US-10-600-070-190 | Sequence 190, App |
| 32 | 409.5 | 10.1 | 205 | 4 | US-10-767-701-41038 | Sequence 41038, A |
| 33 | 401 | 9.9 | 716 | 4 | US-10-600-070-160 | Sequence 160, App |
| 34 | 366 | 9.0 | 99 | 4 | US-10-600-070-124 | Sequence 124, App |
| 35 | 359 | 8.8 | 566 | 4 | US-10-600-070-119 | Sequence 119, App |
| 36 | 344.5 | 8.5 | 702 | 4 | US-10-600-070-100 | Sequence 100, App |
| 37 | 339 | 8.3 | 66 | 4 | US-10-600-070-100 | Sequence 100, App |
| 38 | 328.5 | 8.1 | 573 | 4 | US-10-600-070-120 | Sequence 120, App |
| 39 | 327 | 8.0 | 557 | 4 | US-10-600-070-117 | Sequence 117, App |
| 40 | 320 | 7.9 | 661 | 4 | US-10-600-070-158 | Sequence 158, App |
| 41 | 317.5 | 7.8 | 515 | 4 | US-10-600-070-121 | Sequence 121, App |
| 42 | 308.5 | 7.6 | 524 | 4 | US-10-600-070-118 | Sequence 118, App |
| 43 | 288 | 7.1 | 164 | 4 | US-10-437-963-110670 | Sequence 110670, A |
| 44 | 263 | 6.5 | 66 | 4 | US-10-600-070-99 | Sequence 99, App1 |
| 45 | 257 | 6.3 | 66 | 4 | US-10-600-070-98 | Sequence 98, App1 |

ALIGNMENTS

RESULT 1

US-10-600-070-2

; Sequence 2, Application US/106000070

; Publication No. US20040139500A1

; GENERAL INFORMATION:

; APPLICANT: Osteryoung, Katherine W.

; APPLICANT: Vitha, Stanislav

; APPLICANT: Koksharova, Olga A.

; TITLE OF INVENTION: Plastid Division and Related Genes and Proteins, and Methods of

; TITLE OF INVENTION: Use

; FILE REFERENCE: MSU-08153

; CURRENT APPLICATION NUMBER: US/10/600,070

; CURRENT FILING DATE: 2003-06-20

; NUMBER OF SEQ ID NOS: 206

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 2

; LENGTH: 801

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

; US-10-600-070-2

Query Match 100.0%; Score 4063; DB 4; Length 801;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 801; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|-----|----|--|-----|
| 1 | Qy | MEALSHVIGLSPFQLCRLPPATTKLRSHNTSTTICSASKWADRLSDFNFTSDSSSS | 60 |
| 1 | Db | MEALSHVIGLSPFQLCRLPPATTKLRSHNTSTTICSASKWADRLSDFNFTSDSSSS | 60 |
| 61 | Qy | FATATTATLVSPSIDRPERHVPIPIDFYQVLGAQTHFLTDGIRRAFEARVSKPPQFG | 120 |
| 61 | Db | FATATTATLVSPSIDRPERHVPIPIDFYQVLGAQTHFLTDGIRRAFEARVSKPPQFG | 120 |
| 121 | Qy | FSDDALISRQIIQAACETLSNPRSRREYNEGLLDDDEATVITDVPMDKVPGLCVLQEG | 180 |
| 121 | Db | FSDDALISRQIIQAACETLSNPRSRREYNEGLLDDDEATVITDVPMDKVPGLCVLQEG | 180 |
| 181 | Qy | GETEIVLRVGEALLKERLPKSFQDQVVLVMAFLDVSROMALDPPDFITGYEFVEEAL | 240 |
| 181 | Db | GETEIVLRVGEALLKERLPKSFQDQVVLVMAFLDVSROMALDPPDFITGYEFVEEAL | 240 |
| 241 | Qy | KLQEGGASSLPDLRAQIDETLEEITPRYVLELLGLPLGDDYAAKRLNGLSGVRNIIWS | 300 |
| 241 | Db | KLQEGGASSLPDLRAQIDETLEEITPRYVLELLGLPLGDDYAAKRLNGLSGVRNIIWS | 300 |
| 301 | Qy | VGGGGASALVGGITREKFMNEAFLRMTAAEQVDLFPVATPSNIPAESFVYEVALVAQA | 360 |
| 301 | Db | VGGGGASALVGGITREKFMNEAFLRMTAAEQVDLFPVATPSNIPAESFVYEVALVAQA | 360 |

```
QY 361 FIGKPHLLQADAKQFOOQOAKVAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVIDE 420
Db 361 FIGKPHLLQADAKQFOOQOAKVAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVIDE 420
QY 421 CRMWGLDSDSDSOYRNPAIVEFVLENSNRDNDLPGCLKLLETWLAGVVFPRFRTDKK 480
Db 421 CRMWGLDSDSDSOYRNPAIVEFVLENSNRDNDLPGCLKLLETWLAGVVFPRFRTDKK 480
QY 481 KFKLGDDYDDPMVLSYLERVEVVGQSPAAAAATWARI GAHVKASAMQALQKVPFPRSYTD 540
Db 481 KFKLGDDYDDPMVLSYLERVEVVGQSPAAAAATWARI GAHVKASAMQALQKVPFPRSYTD 540
QY 541 RNSAEPKDVQETVFSVDPVGNVGRDCEPGVFIAEAVRPSNFETNDYAI RAGVSESSVD 600
Db 541 RNSAEPKDVQETVFSVDPVGNVGRDCEPGVFIAEAVRPSNFETNDYAI RAGVSESSVD 600
QY 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFQKYLKSSSFQKQKDMVSSMESDVATI 660
Db 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFQKYLKSSSFQKQKDMVSSMESDVATI 660
QY 721 ETAQLGLVYDYLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
Db 721 ETAQLGLVYDYLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
QY 781 RYEVFWSKSGWKITEGSVLAS 801
Db 781 RYEVFWSKSGWKITEGSVLAS 801

RESULT 2
US-10-600-070-127
; Sequence 127, Application US/106000070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Oosteryoung, Katherine W.
; APPLICANT: Vitha, Stanislav
; APPLICANT: Koksharova, Olga A.
; APPLICANT: Gao, Hong
; TITLE OF INVENTION: Plastid Division and Related Genes and Proteins, and Methods of
; FILE REFERENCE: MSU-08153
; CURRENT APPLICATION NUMBER: US/10/600,070
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 127
; LENGTH: 801
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-600-070-127

Query Match 99.7%; Score 4051; DB 4; Length 801;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MEALSHVIGLSPQLCRLPPATTKLRSHNTSTTICSASKWADRLLSDFNFTSDSSSS 60
Db 1 MEALSHVIGLSPQLCRLPPATTKLRSHNTSTTICSASKWADRLLSDFNFTSDSSSS 60
QY 61 FATATTATLVSPPSIDRPERHVPIIDFYVLGAQTHFLTDGIRAFEARVSKPPQFG 120
Db 61 FATATTATLVSPPSIDRPERHVPIIDFYVLGAQTHFLTDGIRAFEARVSKPPQFG 120
QY 121 FSDALISRQILQAAACETLSNPRSRREYNELLDDEATVITDVPMDKVPKALCVLQEG 180
Db 121 FSDALISRQILQAAACETLSNPRSRREYNELLDDEATVITDVPMDKVPKALCVLQEG 180
QY 181 GETEIVLRVGEALLKERLPKSFQDQVVLVMAALAFLDVSRDAMALDPDFITGYEFVEAL 240
Db 181 GETEIVLRVGEALLKERLPKSFQDQVVLVMAALAFLDVSRDAMALDPDFITGYEFVEAL 240
```

```
Db 181 GETEIVLRVGEALLKERLPKSFQDQVVLVMAALAFLDVSRDAMALDPDFITGYEFVEAL 240
QY 241 KLOEQEGASSIAPDLRAQIDETLEEITPRVYLELLGLPLGDDYAAKRLNGLSGVNRILWS 300
Db 241 KLOEQEGASSIAPDLRAQIDETLEEITPRVYLELLGLPLGDDYAAKRLNGLSGVNRILWS 300
QY 301 VGGGASALVGGLTREKFNHEAFLRMTAAEQVDLFVATPSNIPAESFEVVEVALALVAQA 360
Db 301 VGGGASALVGGLTREKFNHEAFLRMTAAEQVDLFVATPSNIPAESFEVVEVALALVAQA 360
QY 361 FIGKPHLLQADAKQFOOQOAKVAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVIDE 420
Db 361 FIGKPHLLQADAKQFOOQOAKVAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVIDE 420
QY 421 CRMWGLDSDSDSOYRNPAIVEFVLENSNRDNDLPGCLKLLETWLAGVVFPRFRTDKK 480
Db 421 CRMWGLDSDSDSOYRNPAIVEFVLENSNRDNDLPGCLKLLETWLAGVVFPRFRTDKK 480
QY 481 KFKLGDDYDDPMVLSYLERVEVVGQSPAAAAATWARI GAHVKASAMQALQKVPFPRSYTD 540
Db 481 KFKLGDDYDDPMVLSYLERVEVVGQSPAAAAATWARI GAHVKASAMQALQKVPFPRSYTD 540
QY 541 RNSAEPKDVQETVFSVDPVGNVGRDCEPGVFIAEAVRPSNFETNDYAI RAGVSESSVD 600
Db 541 RNSAEPKDVQETVFSVDPVGNVGRDCEPGVFIAEAVRPSNFETNDYAI RAGVSESSVD 600
QY 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFQKYLKSSSFQKQKDMVSSMESDVATI 660
Db 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFQKYLKSSSFQKQKDMVSSMESDVATI 660
QY 661 GSVRADDSEALPRMDARTAEINISWKQIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAA 720
Db 661 GSVRADDSEALPRMDARTAEINISWKQIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAA 720
QY 721 ETAQLGLVYDYLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
Db 721 ETAQLGLVYDYLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
QY 781 RYEVFWSKSGWKITEGSVLAS 801
Db 781 RYEVFWSKSGWKITEGSVLAS 801

RESULT 3
US-10-600-070-129
; Sequence 129, Application US/106000070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Oosteryoung, Katherine W.
; APPLICANT: Vitha, Stanislav
; APPLICANT: Koksharova, Olga A.
; APPLICANT: Gao, Hong
; TITLE OF INVENTION: Plastid Division and Related Genes and Proteins, and Methods of
; FILE REFERENCE: MSU-08153
; CURRENT APPLICATION NUMBER: US/10/600,070
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 129
; LENGTH: 801
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-600-070-129

Query Match 99.7%; Score 4051; DB 4; Length 801;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MEALSHVIGLSPQLCRLPPATTKLRSHNTSTTICSASKWADRLLSDFNFTSDSSSS 60
Db 1 MEALSHVIGLSPQLCRLPPATTKLRSHNTSTTICSASKWADRLLSDFNFTSDSSSS 60
```

QY 61 FATATTATLVSPPSIDRPERHVPIDIDFYQVLGAOTHFLTDGIRRAFEARVSKPPQFG 120
DB 61 FATATTATLVSPPSIDRPERHVPIDIDFYQVLGAOTHFLTDGIRRAFEARVSKPPQFG 120
QY 121 FSDALLSRRQILQAACETLSNPSRREYNEGLDDEEATVITDVPWMDKVPKALCVLQEG 180
DB 121 FSDALLSRRQILQAACETLSNPSRREYNEGLDDEEATVITDVPWMDKVPKALCVLQEG 180
QY 181 GETEIVLRVGEALLKERLPKSFQDQVVLWALAFIDVSRDAMALDPPDFITGYEFVEEAL 240
DB 181 GETEIVLRVGEALLKERLPKSFQDQVVLWALAFIDVSRDAMALDPPDFITGYEFVEEAL 240
QY 241 KLLQEBGASSLAPDLRAQIDETLEETPRYVLELGLPLGDDYAAKRLNGLSGVRNLIWS 300
DB 241 KLLQEBGASSLAPDLRAQIDETLEETPRYVLELGLPLGDDYAAKRLNGLSGVRNLIWS 300
QY 301 VGGGASALVGGLTREKFMNEAFLRMTAAEQVDL FVATPSNIPAESFEVVEVALALVAQA 360
DB 301 VGGGASALVGGLTREKFMNEAFLRMTAAEQVDL FVATPSNIPAESFEVVEVALALVAQA 360
QY 361 FIGKPHLLQADKQFOOQOAKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
DB 361 FIGKPHLLQADKQFOOQOAKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
QY 421 CRMWLGDSQYRNPAIVEFVLENSNRDNDLPGCLKLETWLAGVPPFRDRTDK 480
DB 421 CRMWLGDSQYRNPAIVEFVLENSNRDNDLPGCLKLETWLAGVPPFRDRTDK 480
QY 481 KFKLGDYDDPWLVSYLEVERVEVWGSPPLAAATWARI GAHVKSAMQALQKVPFSRYTD 540
DB 481 KFKLGDYDDPWLVSYLEVERVEVWGSPPLAAATWARI GAHVKSAMQALQKVPFSRYTD 540
QY 541 RNSAEPKDVQETVFSVDPVGNVGRDGEVGFIAEAVRPSNFETNDYAIRAGVSESSVD 600
DB 541 RNSAEPKDVQETVFSVDPVGNVGRDGEVGFIAEAVRPSNFETNDYAIRAGVSESSVD 600
QY 601 ETTVEMSVADMLKEASVKILAAAGVAIGLISLFSOKYFLKSSSSQKQKMDVSSMESDVATI 660
DB 601 ETTVEMSVADMLKEASVKILAAAGVAIGLISLFSOKYFLKSSSSQKQKMDVSSMESDVATI 660
QY 661 GSVRADDSALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPVLDGRMLKIWTDRAA 720
DB 661 GSVRADDSALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPVLDGRMLKIWTDRAA 720
QY 721 ETAQLGLVYDYTLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
DB 721 ETAQLGLVYDYTLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
QY 781 RYEVFWSKSGWKITEGSLAS 801
DB 781 RYEVFWSKSGWKITEGSLAS 801

RESULT 4

US-10-600-070-131
; Sequence 131, Application US/106000070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Osteryoung, Katherine W.
; APPLICANT: Vitha, Stanislaw
; APPLICANT: Koksharova, Olga A.
; APPLICANT: Gao, Hongbo
; TITLE OF INVENTION: Plastid Division and Related Genes and Proteins, and Methods of
; TITLE OF INVENTION: Use
; FILE REFERENCE: MSU-08153
; CURRENT APPLICATION NUMBER: US/10/600,070
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 131
; LENGTH: 801
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana

US-10-600-070-131

Query Match 99.7%; Score 4051; DB 4; Length 801;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MEALSHVIGIGLSPQLCRLPPATTKLRSHNTSTTICSAKWADRLLDNFNTSSSSSS 60
DB 1 MEALSHVIGIGLSPQLCRLPPATTKLRSHNTSTTICSAKWADRLLDNFNTSSSSSS 60
QY 61 FATATTATLVSPPSIDRPERHVPIDIDFYQVLGAOTHFLTDGIRRAFEARVSKPPQFG 120
DB 61 FATATTATLVSPPSIDRPERHVPIDIDFYQVLGAOTHFLTDGIRRAFEARVSKPPQFG 120
QY 121 FSDALLSRRQILQAACETLSNPSRREYNEGLDDEEATVITDVPWMDKVPKALCVLQEG 180
DB 121 FSDALLSRRQILQAACETLSNPSRREYNEGLDDEEATVITDVPWMDKVPKALCVLQEG 180
QY 181 GETEIVLRVGEALLKERLPKSFQDQVVLWALAFIDVSRDAMALDPPDFITGYEFVEEAL 240
DB 181 GETEIVLRVGEALLKERLPKSFQDQVVLWALAFIDVSRDAMALDPPDFITGYEFVEEAL 240
QY 241 KLLQEBGASSLAPDLRAQIDETLEETPRYVLELGLPLGDDYAAKRLNGLSGVRNLIWS 300
DB 241 KLLQEBGASSLAPDLRAQIDETLEETPRYVLELGLPLGDDYAAKRLNGLSGVRNLIWS 300
QY 301 VGGGASALVGGLTREKFMNEAFLRMTAAEQVDL FVATPSNIPAESFEVVEVALALVAQA 360
DB 301 VGGGASALVGGLTREKFMNEAFLRMTAAEQVDL FVATPSNIPAESFEVVEVALALVAQA 360
QY 361 FIGKPHLLQADKQFOOQOAKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
DB 361 FIGKPHLLQADKQFOOQOAKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
QY 421 CRMWLGDSQYRNPAIVEFVLENSNRDNDLPGCLKLETWLAGVPPFRDRTDK 480
DB 421 CRMWLGDSQYRNPAIVEFVLENSNRDNDLPGCLKLETWLAGVPPFRDRTDK 480
QY 481 KFKLGDYDDPWLVSYLEVERVEVWGSPPLAAATWARI GAHVKSAMQALQKVPFSRYTD 540
DB 481 KFKLGDYDDPWLVSYLEVERVEVWGSPPLAAATWARI GAHVKSAMQALQKVPFSRYTD 540
QY 541 RNSAEPKDVQETVFSVDPVGNVGRDGEVGFIAEAVRPSNFETNDYAIRAGVSESSVD 600
DB 541 RNSAEPKDVQETVFSVDPVGNVGRDGEVGFIAEAVRPSNFETNDYAIRAGVSESSVD 600
QY 601 ETTVEMSVADMLKEASVKILAAAGVAIGLISLFSOKYFLKSSSSQKQKMDVSSMESDVATI 660
DB 601 ETTVEMSVADMLKEASVKILAAAGVAIGLISLFSOKYFLKSSSSQKQKMDVSSMESDVATI 660
QY 661 GSVRADDSALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPVLDGRMLKIWTDRAA 720
DB 661 GSVRADDSALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPVLDGRMLKIWTDRAA 720
QY 721 ETAQLGLVYDYTLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
DB 721 ETAQLGLVYDYTLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
QY 781 RYEVFWSKSGWKITEGSLAS 801
DB 781 RYEVFWSKSGWKITEGSLAS 801

RESULT 5

US-10-739-930-5771
; Sequence 5771, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930

; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 5771
; LENGTH: 801
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C13643_1.p
US-10-739-930-5771

Query Match 99.7%; Score 4051; DB 5; Length 801;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MEALSHVGISLSPQICRLPPATTKLRRSHNTTTCASAKWADRLSDNFNTSSSSS 60
DB 1 MEALSHVGISLSPQICRLPPATTKLRRSHNTTTCASAKWADRLSDNFNTSSSSS 60

QY 61 FATATTATLVSIPPSIDRPERHVPIPIDEFYVLGAQTHFLTDGIRAFEARVSKPPQF 120
DB 61 FATATTATLVSIPPSIDRPERHVPIPIDEFYVLGAQTHFLTDGIRAFEARVSKPPQF 120

QY 121 FSDALISRRQIIQAACETLSNPRSRREYNEGLDDEEATVITDVPWDKVPKALCVLQEG 180
DB 121 FSDALISRRQIIQAACETLSNPRSRREYNEGLDDEEATVITDVPWDKVPKALCVLQEG 180

QY 181 GETEIVLRVGEALLKERLPKSKFQDVVLMALAFDVSRDAMALDPPDFITGYEFVEEAL 240
DB 181 GETEIVLRVGEALLKERLPKSKFQDVVLMALAFDVSRDAMALDPPDFITGYEFVEEAL 240

QY 241 KLLQEGASLAPDLRAQIDETLEEITPRYVLELGLPLGDDYAAKRLNGLSGVRNILWS 300
DB 241 KLLQEGASLAPDLRAQIDETLEEITPRYVLELGLPLGDDYAAKRLNGLSGVRNILWS 300

QY 301 VGGGGASALVGGITREKFMNEAFRTAAEQVDLFVATPSNIPAESFEVVEVALALVAQA 360
DB 301 VGGGGASALVGGITREKFMNEAFRTAAEQVDLFVATPSNIPAESFEVVEVALALVAQA 360

QY 361 FIGKPHLLQADAKQFQOLQAQVAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
DB 361 FIGKPHLLQADAKQFQOLQAQVAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420

QY 421 CRMWGLDSDSQRNPAIVEFVLENSNRDNDLPLGCKLLETWLAGVVPFRDTKOK 480
DB 421 CRMWGLDSDSQRNPAIVEFVLENSNRDNDLPLGCKLLETWLAGVVPFRDTKOK 480

QY 481 KFKLGYYDDPMVLSYLERVEVVGSPPLAAATWARIAGHVKASAMQALOKVFPBSRYTD 540
DB 481 KFKLGYYDDPMVLSYLERVEVVGSPPLAAATWARIAGHVKASAMQALOKVFPBSRYTD 540

QY 541 RNSAEPKDVQETVFSVDPVGNVGRDCEPGVFTAEAVRPSNFETNDYAIRAGVSESSVD 600
DB 541 RNSAEPKDVQETVFSVDPVGNVGRDCEPGVFTAEAVRPSNFETNDYAIRAGVSESSVD 600

QY 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFSQKYLKSSSSSFQKDMVSSMESDVATI 660
DB 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFSQKYLKSSSSSFQKDMVSSMESDVATI 660

QY 661 GSVRADSEALPRMDARTAEINISVKQIKSLAFGPDHRIEMLPVLDGRMLKIWTDRAA 720
DB 661 GSVRADSEALPRMDARTAEINISVKQIKSLAFGPDHRIEMLPVLDGRMLKIWTDRAA 720

QY 721 ETAQLGLVYDYLKLSVDSVTSADGTRALVEATLEESACLSDLVHPENNAIDVRYTTT 780
DB 721 ETAQLGLVYDYLKLSVDSVTSADGTRALVEATLEESACLSDLVHPENNAIDVRYTTT 780

QY 781 RYEVFWKSGWKITEGSLVAS 801
DB 781 RYEVFWKSGWKITEGSLVAS 801

RESULT 6
US-10-600-070-123

; Sequence 123, Application US/10600070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Osteryoung, Katherine W.
; APPLICANT: Vitha, Stanislav
; APPLICANT: Koksharova, Olga A.
; APPLICANT: Gao, Hongbo
; TITLE OF INVENTION: Plastid Division and Related Genes and Proteins, and Methods of
; FILE OF INVENTION: Use
; CURRENT APPLICATION NUMBER: US/10/600,070
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 123
; LENGTH: 578
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-600-070-123

Query Match 72.0%; Score 2926; DB 4; Length 578;
Best Local Similarity 99.8%; Pred. No. 3.1e-237;
Matches 577; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 79 RPERHVPIPIDEFYVLGAQTHFLTDGIRAFEARVSKPPQFSGFSDDALISRRQIIQAACE 138
DB 1 RPERHVPIPIDEFYVLGAQTHFLTDGIRAFEARVSKPPQFSGFSDDALISRRQIIQAACE 60

QY 139 TLSNPSRRRYNEGLDDEEATVITDVPWDKVPKALCVLQEGTEIVLRVGEALLKERL 198
DB 61 TLSNPSRRRYNEGLDDEEATVITDVPWDKVPKALCVLQEGTEIVLRVGEALLKERL 120

QY 199 PKSFQDVVLMALAFDVSRDAMALDPPDFITGYEFVEEALKLLQEGASSLAPDLRAQ 258
DB 121 PKSFQDVVLMALAFDVSRDAMALDPPDFITGYEFVEEALKLLQEGASSLAPDLRAQ 180

QY 259 IDTLEETIPRYVLELGLPLGDDYAAKRLNGLSGVRNILWSVGGGASALVGGITREKF 318
DB 181 IDTLEETIPRYVLELGLPLGDDYAAKRLNGLSGVRNILWSVGGGASALVGGITREKF 240

QY 319 MNEAFRTAAEQVDLFVATPSNIPAESFEVVEVALALVAQAFIGKPHLLQADAKQFQ 378
DB 241 MNEAFRTAAEQVDLFVATPSNIPAESFEVVEVALALVAQAFIGKPHLLQADAKQFQ 300

QY 379 LQAQKVAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVDCEMRWGLDSDSQRNPA 438
DB 301 LQAQKVAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVDCEMRWGLDSDSQRNPA 360

QY 439 IVEFVLENSNRDNDLPLGCKLLETWLAGVVPFRDTKOKKFKLGYYDDPMVLSYLE 498
DB 361 IVEFVLENSNRDNDLPLGCKLLETWLAGVVPFRDTKOKKFKLGYYDDPMVLSYLE 420

QY 499 RVEVVGSPPLAAATWARIAGHVKASAMQALOKVFPBSRYTDRNSAEPKDVQETVFSVDP 558
DB 421 RVEVVGSPPLAAATWARIAGHVKASAMQALOKVFPBSRYTDRNSAEPKDVQETVFSVDP 480

QY 559 VGNVGRDCEPGVFTAEAVRPSNFETNDYAIRAGVSESSVDYTTVMSVADMLKEASVK 618
DB 481 VGNVGRDCEPGVFTAEAVRPSNFETNDYAIRAGVSESSVDYTTVMSVADMLKEASVK 540

QY 619 ILAAGVAIGLISLFSQKYLKSSSSSFQKDMVSSMESD 656
DB 541 ILAAGVAIGLISLFSQKYLKSSSSSFQKDMVSSMESD 578

RESULT 7
US-10-600-070-125
; Sequence 125, Application US/10600070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Osteryoung, Katherine W.
; APPLICANT: Vitha, Stanislav
; APPLICANT: Koksharova, Olga A.

US-10-600-070-122

```
Query Match          32.8%; Score 1332; DB 4; Length 525;
Best Local Similarity 48.8%; Pred. No. 5.9e-103;
Matches 287; Conservative 88; Mismatches 137; Indels 76; Gaps 11;

QY 81 ERHVPIPIIDFYVLGAQTHFLTDGIRRAFEARVSKPPQFGSDDALISROILQAACETL 140
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 3 ERSPLQVDFYVLGAEPHFLGDIRRAFEARVSKPPQFGSDDALISROILQAACETL 62
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 141 SNPRSRREYNEGLDDEEATVITDVPWQVPGALCVLQEGGTEIVLRVGEALLKERLPK 200
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 63 MQNRSRTQDRALSENREAEATWTDWCK-----EAGEALAVLVGEGQLLDRPK 113

QY 201 SFQDVVLVMAALFVDSRDAMALDPDFTITGVFEVEALKLLQEGCASLAPDLRAQID 260
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 114 RFQDVVLAMALAYVLDLSRDMAASPDPVIGCCEVLERALKLLQEGCASLAPDLSSQID 173

QY 261 ETLERITPRVYLELGLPGDDYAAKRLNGLSGVRNILMSVGGGASALVGLTREKFN 320
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 174 ETLERITPRVYLELGLPGDDYAAKRLNGLSGVRNILMSVGGGASALVGLTREKFN 233

QY 321 EAPLRTAAEQVDLFVATPSNIPAESFEVEVALALVAQAFIKKPHLLQDADKQFOOLQ 380
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 234 EAPLRTAAEQVDLFVATPSNIPAESFEVEVALALVAQAFIKKPHLLQDADKQFOOLQ 293

QY 381 QAKVMAEIPAMLYDTRNWEIDFGLERGLCALLIGKVDCECRMWLGDSQYRNPAIV 440
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 294 KFNIGS----HYAYDN---EMDLALERAFCSLLVGDVSKCRMWLGDSQYRNPAIV 345

QY 441 EFVLENSN-RDNDLGLCKLLETWLAGVPRFRDTRKFKLGDYDDPWLSYLER 499
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 346 EFVLENSN-RDNDLGLCKLLETWLAGVPRFRDTRKFKLGDYDDPWLSYLER 405

QY 500 VEVVQGSPLAARATWARICAE-----HYKASAMQALQVFP-SRYTDRNSAEKDVQET 552
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 406 MEGGAGASHLAAALAAKGAQAALGTVKNSAIOAFNKFVPLIEQLDRSAMENT----- 460

QY 553 VFSVDPVGNVGRDGPVGFIAEAVRPSNFETNDYAIRAGVSESSVDETTVMMSVADML 612
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 461 -----KDG-PGGYL-----ENFDQ-----ENAPAHDS 481

QY 613 KEASVKILAAGVAIGLISLFSOKYFLKSSSFQKMDVMSMESDVATI 660
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 482 RNAALKIISAGALFALLAVIGAKY-----LPRKRPLSAIRSEHGSV 522

RESULT 10
US-10-424-599-271849
; Sequence 271849, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 271849
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_87499C.1.pap
US-10-424-599-271849
```

```
Query Match          27.9%; Score 1134; DB 4; Length 344;
Best Local Similarity 66.4%; Pred. No. 1.5e-86;
Matches 223; Conservative 48; Mismatches 53; Indels 12; Gaps 3;
```

```
QY 16 LCKLPPATT-----KLRRS--HNTSTTICSASKWADRLLSDFNFTSDSSSSSFATATT 66
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 12 LCTPHPTTTHPTFKPNKLLRSSLSRGAASLSATSQWAERLIADPQFLGDAAS---TSTS 68
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 67 TATLVSLPPSIDRPERHVPIPIIDFYVLGAQTHFLTDGIRRAFEARVSKPPQFGSDDAL 126
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 69 TUSPSSVPRLDDPERVYSIPLDYLRIILGAEPHFLGDIRRAFEARVSKPPQYAFSNDAL 128
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 127 ISRRQILQAACETLSNPRSRREYNEGLDDEEATVITDVPWQVPGALCVLQEGGTEIV 186
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 129 ISRRQILQAACETLSNPRSRREYNEGLDDEEATVITDVPWQVPGALCVLQEGGTEIV 188
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 187 LRVGEALLKERLPKSFQDVVLVMAALFVDSRDAMALDPDFTITGVFEVEALKLQEE 246
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 189 LEIQGQLLRERLPKTFKQDVVLAMALAFVDSRDAMALSPDPFIAACEMLERALKLQEE 248
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 247 GASSLAPDLRAQIDETLEETITPRVYLELGLPGDDYAAKRLNGLSGVRNILMSVGGGA 306
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 249 GATSLAPDLRAQIDETLEETITPRVYLELGLPGDDYAAKRLNGLSGVRNILMSVGGGA 308
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 307 SALVGLGTRKFMNEAFRLMTAAEQVDLFVATPSNI 342
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 309 ATTAGFTREDFNEAFRLMTAAEQVDLFVATPSNI 344
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 11
US-10-425-115-266516
; Sequence 266516, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 266516
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_174665C.1.pap
US-10-425-115-266516
```

```
Query Match          23.9%; Score 973; DB 4; Length 364;
Best Local Similarity 54.8%; Pred. No. 5.9e-73;
Matches 198; Conservative 54; Mismatches 79; Indels 30; Gaps 4;
```

```
QY 12 SPQCLRLPPATTKLRRSHNT-----STTICSASKWADRLLSDFNFTSDSS 57
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 14 APPAFSPLP-----LRRSHRPPPGPSTCRASRWADRLFPADFHLLPAAADPPAAASSS 68
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 58 SSSFATATTATLVSLPPSIDRPERHVPIPIIDFYVLGAQTHFLTDGIRRAFEARVSKPP 117
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 69 SSPF-----VPIFPEAADRALPLFVDYFKILGAEPHFLGDIRAFESRIAKPP 117
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 118 QGFGSDDALISRRQILQAACETLSNPRSRREYNEGLDDEEATVITDVPWQVPGALCVL 177
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 118 QYGYSTEALVGRQMLQIAHDTLTNOSRTEYDRLSEDRDALTDVDMVKVGVLCVL 177
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 178 QEGGTEIVLRVGEALLKERLPKSFQDVVLVMAALFVDSRDAMALDPDFTITGVFEVE 237
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 178 QEAGEAQLVATGEHLLQDRPKRFQDVVLAMALAYVDISRDMAAASPDPVICCEVLE 237
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 238 EALKLQEGCASLAPDLRAQIDETLEETITPRVYLELGLPGDDYAAKRLNGLSGVRNI 297
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 238 RALKLQEGCASLAPDLRAQIDETLEETITPRVYLELGLPGDDYAAKRLNGLSGVRNI 297
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
```


RESULT 14

| | | |
|---------------------------|---------------------------|-----|
| Best Local Similarity | 40.7%; Pred. No. 5.3e-63; | |
| Matches 191: Conservative | 82; Mismatches 126; | |
| Indels | 70; Gaps | 11; |

[illegible]

```

US-10-600-070-167
; Sequence 167, Application US/106000070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Osteryoung, Katherine W.
; APPLICANT: Vitha, Stanislav
; APPLICANT: Koksharova, Olga A.
; APPLICANT: Gao, Hongo
; TITLE OF INVENTION: Plasmid Division and Related Genes and Proteins, and Methods of
; FILE REFERENCE: MSU-08153
; CURRENT APPLICATION NUMBER: US/10/600,070
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 167
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Nostoc punctiforme
US-10-600-070-167

Query Match          12.9%; Score 524.5; DB 4; Length 768;
Best Local Similarity 24.9%; Pred. No. 1.2e-34;
Matches 212; Conservative 128; Mismatches 286; Indels 227; Gaps 28;

Qy      84  VPIDIFVQLGAQTHFTDGIIRAFEARVSKPPQFGFSDDALISRQILQAACETLSNP 143
Db      1  VRIPLDYRIILGLPLAASEEOLRQAYSRIVOLPRREYSQAAISRKRQLEEEAYVVLSDP 60

Qy      144 RSREYNENGL-----DDEATVIT-----DVPWDKVPGLCVLQ 178
Db      61  QRSTYDQLYLAHYDPPNLAATAVAQENRTESTKRGSTQSLGIRITQDELGALLILQ 120

Qy      179 EGGSTEIVLRVGEALL--KERLPKSFQ-----DVLVLMALAEFDVSRD- 220
Db      121 ELGEVELVILGRPYLVNKNKSATSRKSNLDADEEIEYSAEHPDVVLTVALACLELGREG 180

Qy      221 -----ANALPPDPFITGYEFVEALKLLOEGASSIADPLDRAQIDETLEEITPRYV 271
Db      181 WQOQHYENAATLSL-----TQGE-----LLVREG---LFSSIQAEIQADLYKLRPYRI 225

Qy      272 LELLGLPLGDDYAAKRLNGLSGVRNILMSVGG--GGASALVGGLTRE---KFMNEAFLRMT 327
Db      226 LELLALP--QEKTAERSQGLLEQLNLLDREGGIDGTNDESGNLNIDDFLRFIQOLRNHLT 283

Qy      328 AAQVDLVFATPSNIPAESFEYVEVALALVAQAQFGKPKHLLQADADKQFQQLQAQKVMAM 387
Db      284 VAEQHKLFQAQSKR--SSAVATYLAAYALVARGFAQRQAPALINQARQMLVRLGKRQ---- 337

Qy      388 EIPAWLYTRNNWIDFGLRGLCALLIGKVDCEKMWGLGDSDSOYRNPATVEFVLENS 447
Db      338 -----DVHLRQSLCALLIGQTEATRVLELSQE---YE---ALAFIREKS 376

Qy      448 NRDDNDLPLGLCKLLETWLAGVVFPRFRDTKDKFKLGDYDDPMLVSLYLERVEVWQGP 507
Db      377 -QDSFDLLPGLCLYAEQWLQHEVVFPHFRDLANQQAFLKDYFANQQVQAYLE----- 426

Qy      508 LAAATWARI GAEHVKASAMQALQKVFPSRYTDRNSAEPK-----DVOETVFSV 556
Db      427 --ALPTAQTTNEWAVINPQYFQAKAKNTHPHNNSTKTSASFNHSRVPNPDLPET----- 480

Qy      557 DPVGNVGRDGEPGVFAEAVRPSNFENDYAIRAGVSESSVDVETTVMSVADMLK--- 613
Db      481 -----PTK--ETSEYFNFSPPMWSSSGSIKSEVPAAERMSRGT 516

Qy      614 ----EASVKILAAG----- 623
Db      517 NQHLNGSAKSAASHGNQKRRRKPTPSASRERI PDNRPHSRPRRRRTFTANTIEGTRLV 576

Qy      624 --VAIGLISL-----FSQYFLKSSSQFKDKMWSMESDVIATIGSVRADDS 669
Db      577 WRVFISLWSILVFWLATTTFGLWKNLFFPQPSPDLQLFVQINQPPPLPIPDNPKPDESE 636

```

```

Qy 670 ALPRMDARTAEINISKWKIKSLAFGPDHRIEMLPEVL DGRMLKIWTDRAAETAQLGLVY 729
Db   :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 637 EGPLTNAB-AEEVIHTLSTKAAALGNHEINNLEQILTGSALSQR-LIAQQNKLDNRY 694
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Qy 730 ---DYTLKLKSVDSVTVADGTFRALVEATLESACISDLVHPENNATDVRTTYTRYEVFW 786
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 695 RKDFHLSLKIESVEKIGLFAD--RAAVEATVKEVDTLYENNQFNSSND--KLRVRDYDLIR 750
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Qy 787 SKGWKKITEGSVL 799
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 751 ERGKWRIQTSTVV 763
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 15
US-10-424-599-177901
; Sequence 177901, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kowalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 177901
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_131660C.1.pep
US-10-424-599-177901

Query Match          12.3%; Score 499; DB 4; Length 204;
Best Local Similarity 51.2%; Pred.No.2.1e-33;
Matches 104; Conservative 39; Mismatches 58; Indels 2; Gaps 2;

Qy 601 ETTVMYSVADM LKEASVKILAAGVAIGLISFLFSQKYF-LKSSSSFQFKDMVSMSESVDAT 659
Db    2 ETRKGVFITEBKHASVOLIMCAVGIVGLVTLVGLKPLPTRNGSPILRKMTGSA MVSDTIN 61
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Qy 660 IGSVRADD-S-EALPMDARTAEINISKWKIKSLAFGPDHRIEMLPEVL DGRMLKIWTDR 718
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 62 IGSLGDEEKVEQLPKMDARVAEALVRKMOSVKSEAFGPDC HLGRLHEVL DGEMLKIWTDR 121
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Qy 719 AAETAOLGLVVDYTLLKLSVDSVTVSADCTRALVEATLTLESACISDLVHPNNNATDVRTY 778
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 122 AAEAERGSWDYDTLEDLNIDSVTI SQNRRAVVETTLKESTHLNAVCHPOHDASNRPY 181
      :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Qy 779 TTRYEVFWSKSGWKITEGSVLAS 801
Db 182 TTRYEMSETGAEWKIVEGVAVLES 204

```

Search completed: December 4, 2005, 12:42:30
Job time : 170 secs

| Result No. | Score | Query | | Length | DB | ID | Description |
|------------|-------|-------|------|--------|--------------------|--------------------|-------------|
| | | Match | % | | | | |
| 1 | 116 | 2.9 | 1448 | 6 | US-10-485-517-212 | Sequence 212, Appl | |
| 2 | 111.5 | 2.7 | 3704 | 6 | US-10-513-786-1 | Sequence 1, Appli | |
| 3 | 109.5 | 2.7 | 661 | 6 | US-10-467-657-1266 | Sequence 1266, Ap | |
| 4 | 107 | 2.6 | 908 | 6 | US-10-467-657-1070 | Sequence 1070, Ap | |
| 5 | 105.5 | 2.6 | 817 | 6 | US-10-793-626-50 | Sequence 50, Appl | |
| 6 | 105.5 | 2.6 | 817 | 6 | US-10-793-626-1528 | Sequence 1528, Ap | |
| 7 | 104.5 | 2.6 | 402 | 6 | US-10-858-730-30 | Sequence 30, Appl | |
| 8 | 104 | 2.6 | 627 | 6 | US-10-467-657-5432 | Sequence 5432, Ap | |
| 9 | 103.5 | 2.5 | 715 | 6 | US-10-467-657-70 | Sequence 70, Appl | |
| 10 | 103.5 | 2.5 | 715 | 6 | US-10-467-657-3672 | Sequence 3672, Ap | |
| 11 | 103.5 | 2.5 | 2004 | 6 | US-10-467-657-84 | Sequence 84, Appl | |
| 12 | 103.5 | 2.5 | 2004 | 6 | US-10-467-657-6322 | Sequence 6322, Ap | |
| 13 | 103 | 2.5 | 565 | 7 | US-11-074-176-98 | Sequence 98, Appl | |
| 14 | 101 | 2.5 | 403 | 6 | US-10-858-730-27 | Sequence 27, Appl | |
| 15 | 101 | 2.5 | 711 | 6 | US-10-467-657-2966 | Sequence 2966, Ap | |
| 16 | 100.5 | 2.5 | 1663 | 6 | US-10-983-545-6 | Sequence 6, Appli | |
| 17 | 99 | 2.4 | 943 | 6 | US-10-467-657-5508 | Sequence 5508, Ap | |
| 18 | 98 | 2.4 | 611 | 7 | US-11-082-389-436 | Sequence 436, App | |
| 19 | 97.5 | 2.4 | 505 | 6 | US-10-467-657-5148 | Sequence 5148, Ap | |
| 20 | 97.5 | 2.4 | 791 | 6 | US-10-467-657-5014 | Sequence 5014, Ap | |
| 21 | 97.5 | 2.4 | 1192 | 6 | US-10-858-730-72 | Sequence 72, Appl | |
| 22 | 96.5 | 2.4 | 724 | 6 | US-10-793-626-3082 | Sequence 3082, Ap | |
| 23 | 96.5 | 2.4 | 1152 | 7 | US-11-053-822-308 | Sequence 308, App | |
| 24 | 95.5 | 2.4 | 566 | 6 | US-10-467-657-3302 | Sequence 3302, Ap | |
| 25 | 94.5 | 2.3 | 1020 | 6 | US-10-513-786-4 | Sequence 4, Appli | |

| | | | | | | |
|---|-----|--|---|---------------------------|-------------|-----|
| Qy | 306 | ASALV---- | GGITREKFNW---- | EAPLRMTAAEQVDFV---- | ATPSNTPAESF | 347 |
| | | | | | | |
| Db | 268 | AFALTWLBEGSCANESYVNLIPTSLGGTHAEGLKQAVFNANVNFNLHNLPRGVKQVQSD | 327 | | | |
| | | | | | | |
| Qy | 348 | EWY-EVALALVAQAFIGKPHLLQDADKQFOQLQAKVMAME---- | IPAMLVDTRNNW-- | 400 | | |
| | | | | | | |
| Db | 328 | DVFGKTAFLVSARVL----- | DPQFGQTKDKLTNRDALKVAASVGDPLELWLN | 376 | | |
| | | | | | | |
| Qy | 401 | -EIDFG----- | LER----- | GLCALLIGKVDRCRMWLGDSDSQYR | 435 | |
| | | | | | | |
| Db | 377 | QNWDFGKKIAELAIKQARIRSVKIEKKXGSGV-AILPGKLTDC----- | ESEDIREN | 429 | | |
| | | | | | | |
| Qy | 436 | NPAIVEFVLNSNRDDNDLPGLCKLLETW----- | LAGVVPPRFRDTKDKKFKLGDYY | 488 | | |
| | | | | | | |
| Db | 430 | ELFLVE----- | GDSSGSAKLARDKATQAILPLRGKVLNSFEVHPDQLFGNAETH | 479 | | |
| | | | | | | |
| Qy | 489 | D----- | DPMLVSYLERVEVVQGSPLAAATMARIGAHEHVKASAMQALQKVFPSSRYTDR | 541 | | |
| | | | | | | |
| Db | 480 | DISVAVGVDPHAIHNDHPDLSLRGKIAILSD-ADVDGSHIQVLLLTLYRHFPLKLVADG | 538 | | | |
| | | | | | | |
| Qy | 542 | N--SAEPKDVQETVFSVDVPGNVGRDGEPPGVFTAEAVRPSENFETNDYAIIRAGVSESV | 599 | | | |
| | | | | | | |
| Db | 539 | HIYVAQP----- | PLFRVD----- | VNAQGS----- | KPAKF----- | 567 |
| | | | | | | |
| Qy | 600 | DEITVEMSVADMKEASVKILLAAAGVAIGLISLFSQKYFLKSSSSFOROMVSSMESDVAT | 659 | | | |
| | | | | | | |
| Db | 568 | DQNELD-GILERLQKEGVKATAYSI----- | SRFKGLGEMNPDLQKD--TTMHPDTRR | 616 | | |
| | | | | | | |
| Qy | 660 | IGSVRADDSALPRMDARTASNIYSVKWQIKSLAPGDPDHRIEMLPEVLDRMLKIWTORA | 719 | | | |
| | | | | | | |
| Db | 617 | LLQVQIPEG----- | ADDETRDI FVKLMCKGEAA----- | RRAMMERE | 653 | |
| | | | | | | |
| Qy | 720 | AETAQLGL 727 | | | | |
| | | | | | | |
| Db | 654 | GDTAQLDI 661 | | | | |
| | | | | | | |
| RESULT 4 | | | | | | |
| US-10-467-657-1070 | | | | | | |
| ; Sequence 1070, Application US/10467657 | | | | | | |
| ; Publication No. US20050260581A1 | | | | | | |
| ; GENERAL INFORMATION: | | | | | | |
| ; APPLICANT: CHIRON SPA | | | | | | |
| ; APPLICANT: FONTANA Maria Rita | | | | | | |
| ; APPLICANT: PIZZA Mariagrazia | | | | | | |
| ; APPLICANT: MASIGNANI Vega | | | | | | |
| ; APPLICANT: MONACI Elisabetta | | | | | | |
| ; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS | | | | | | |
| ; FILE REFERENCE: | | | | | | |
| ; CURRENT APPLICATION NUMBER: US/10/467,657 | | | | | | |
| ; CURRENT FILING DATE: 2003-08-11 | | | | | | |
| ; PRIOR APPLICATION NUMBER: GB-0103424.8 | | | | | | |
| ; PRIOR FILING DATE: 2001-02-12 | | | | | | |
| ; NUMBER OF SEQ ID NOS: 9218 | | | | | | |
| ; SOFTWARE: SeqWin99, version 1.04 | | | | | | |
| ; SEQ ID NO 1070 | | | | | | |
| ; LENGTH: 908 | | | | | | |
| ; TYPE: PRT | | | | | | |
| ; ORGANISM: Neisseria gonorrhoeae | | | | | | |
| US-10-467-657-1070 | | | | | | |

```

Query Match      2.6%; Score 107; DB 6; Length 908;
Best Local Similarity 19.3%; Pred.No.1.2;
Matches 154; Conservative 108; Mismatches 250; Indels 286; Gaps 37;

Qy    143 PRSRREYNEGLLDDEEAT-----VITDPMDKVPGALCVLQEGGE--TEIVLRVGE----- 191
      |||         |||         |||         |||         |||         |||         |||
Db    62 PDDLRPQAEALPDILRLTGWPLVGVEADDVIGTLA--KQGAEHGLRVIVSTGDKDMA 119

Qy    192 ALLKERLPKFKQDVLVMALAFLDVS-----RDAMAL---DPDPFITGVHFV 236
      |||         |||         |||         |||         |||         |||         |||
Db   120 QLVDERV-----TLVNMTGETLIDSGVKAKFGRPDIIRDYLIALLGDKVDNVPVGEKC 173

```

Query Match

2.6%; Score 105.5; DB 6; Length 817;

7;

length

6:

2

105

505

2.6

Mat

Que

113

GIVEN

OKAY

STAY

1074

CONCLUSIONS

ETIT

OSMT

IT--

-A-
E

75 02

25

| | |
|--|--|
| Best Local Similarity 18.9%; Pred. No. 1.3; Matches 141; Conservative 108; Mismatches 248; Indels 251; Gaps 37; | |
| Qy | 124 DALISRQIIQAAACETLSNPRSRNEGLLDDDEATVITDPMDKVPVGCALCVLQEGEET 183 |
| Db | 180 DPVVGDRDKETRVIEVL-----SRTKNPVLIGE-----PGV-----GKT 215 |
| Qy | 184 EIVLRVGEALLKERLPKSKQDVVVLVMAFLDVSRDAMALDPDFDTGYEF---VEEAL 240 |
| Db | 216 AIAEGLAQAIKNEVPETLK-----DKRWSLDMGTVVAGTKYRGEPEERL 261 |
| Qy | 241 KLIQEE-----GASSLAPDLRAQIDETLEEITPRYVLELGLPLGDD 282 |
| Db | 262 KKVMEETHQAGNVILFIDELHTLVGAGGAGDA--SNILKPALARGELQCIGATTLDE 319 |
| Qy | 283 Y-----AAKR-----LNGLSGRVNLISVGGGASALVGGLTRE 316 |
| Db | 320 YRKNIKDAALERFPQIQVDEPTVEITIELKGLRDYRAHHRINISDRALEAAKLSD 379 |
| Qy | 317 KFMNEAFLRMTAAEQVDFVA-----TPSNIIPAESFEVEYA---LALVAQAFIGK 364 |
| Db | 380 RYVSDFRLPKADILIDEASKVRLKSHSTPSNLKEIQEIDKVKNEKDAAVHAQEF--- 436 |
| Qy | 365 KPHLLDQADKFOOLOQAKWAMEIPAMLVYDTRNNWE-----IDFGL-ERGLCALLIGKV 418 |
| Db | 437 -ENAAANLRDKQSLKQOYE-----DAKNWKNQAQGLDALTALSEENIAEVIAG-- 482 |
| Qy | 419 DECRMWGL-----DSEDQYRNPALVEFVLENSRDNDDDLPCLKILLETWLAGVVPF 472 |
| Db | 483 -----WTGIFLTKINETESDRLLN-----LEDTLHKRVYQONDVNSISKAVRRARAGL-- 531 |
| Qy | 473 RFRDTKDKFKLGDY-YDDPMVLVSLRVEVQGSPLAAATAWARG-AEHVKASAMAQAL 530 |
| Db | 532 -----KDPKRPISGFIFLGTGVGCKTELARALAESMEGEDDAMIRVDMSEFMKHAVSRL 586 |
| Qy | 531 QKVFP-----SRYTDRNSAEPKV-----QETVFSV-----DPVGN 562 |
| Db | 587 VGAPPGVGHDDGGQLTKVRRKPYSVILFDEIEKAHPDVFNILLVLDGHLTDTKGRT 646 |
| Qy | 563 VGRDGEQGVIAEAVRSENSEPNNDYAIRAGVSESSVDETTVMVSADMLKBAKASVKILAA 622 |
| Db | 647 V--DFRNTVITMSTNVGAQELQDQRFAGFGGASEGS-DYETVRKT---MKKE----- 692 |
| Qy | 623 GVALGLISLFSQKYFLAKSSSFQKDWSSNESPVATIGSVRADDDSEALPRMDARTAE 682 |
| Db | 693 -----LKNSP---RPEFLNRVD-DIIVFHKLTDKDELKEIVTM-----M 726 |
| Qy | 683 VSKWQIKISLAFGDPDHRHIEMLPEVLDDGRMLKWIWTDRAAE-TAOLGLVYDVTLLKLSVDVS 741 |
| Db | 727 VNKL-----THRLS-----EQNINIVTDKAKEIAEGYDPEY----- 760 |
| Qy | 742 TVSADGTTRALVEA---TLEESACLSDLV 766 |
| Db | 761 -----GARPLIRAIOKTVEDN--LSELI 781 |

RESULT 6

US-10-793-626-1528
 ; Sequence 1528, Application US/10793626
 ; Publication No. US20050255478A1
 ; GENERAL INFORMATION:
 ; APPLICANT: KIMMERLY, WILLIAM JOHN
 ; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
 ; FILE REFERENCE: P034800S
 ; CURRENT APPLICATION NUMBER: US/10793,626
 ; CURRENT FILING DATE: 2004-03-04
 ; PRIOR APPLICATION NUMBER: 60/164,258
 ; PRIOR FILING DATE: 1999-11-09
 ; NUMBER OF SEQ ID NOS: 4472
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1528
 ; LENGTH: 817
 ; TYPE: PRT

RESULT 7

RESUB 1 1 1
US-10-858-730-30
; Sequence 30, Application US/10858730
; Publication NO. US20050255568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.

```

; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgev, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; TITLE OF INVENTION: PRODUCTION
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 402
; TYPE: PRM
; ORGANISM: Streptomyces coelicolor
US-10-858-730-30

Query Match          2.6%; Score 104.5; DB 6; Length 402;
Best Local Similarity 21.0%; Pred.No. 0.53;
Matches    70; Conservative 50; Mismatches 103; Indels 111; Gaps 14;

QY   94 LGAOTHTLTGIRRAPEARV-----SKPQGFGSDALISRRILQAACETLL 140
      :|:::||:|:||||:|::||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db   96 IGAOSPDIAGVDTYAYENRVEGDDELDROGADGOGLMGFASDETPTLMPLPVFLAHLR 155
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~::~:~::

QY   141 SNPRSREYNELGLDDBEATVTDPWDKVPFGALCVLQEGGETEIVLRV-GEALLKERLP 199
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~::~:~::

Db   156 SKRLSEVRKN-----GTIPYLRPDGKTQTVIYEYDC----- 186
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~::~:~::

QY   200 KSFQDVVLWMAFLDVDSRDMALDDPPDTITGVTFVFEEALKLLOEGCASSLAPDLRAQI 259
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~::~:~::

Db   187 KAVERLDTVVVSSQHASDI--DLESLLAPDI---KEFVVE-----PELKALL 227
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~::~:~::

QY   260 DETLEEITPRVULEL-----LCPLPGDDYYAKKLNGLSGRNLMWSVG-----CGGA- 306
      ::::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~::~:~::

Db   228 EDGIKDITENYRLVANPTGTRFTGGPMGD-----AGLTRKIITIITYGGMARHGGAFF 280
      :|:|:|:|:|:|:|:|:|:|:|:|:|:~::~:~::

QY   307 -----SALVGGTLTREKFWMNEAFLRMTAAEQVDLFVAT--PSNIP 343
      :|:|:|:|:|:|:|:|:|:|:|:|:|:~::~:~::

Db   281 SGKDPKVDRSAAAYMRWVAKVNVVAAGLAARCEVOVAY-AIGKAEPVGLFVETFGTAKVD 339
      :|:|:|:|:|:|:|:|:|:|:~::~:~::

QY   344 AESF-----EVYEVALVALVAQAFIGKKPHLLQDA 372
      :|:|:|:|:|:|:~::~:~::

Db   340 TEKTEKAIDEVFDLRPAAIIRALDLLRFPIYAOTA 373
      :|:|:|:|:|:~::~:~::
```

```

RESULT 8
US-10-467-657-5432
; Sequence 5432, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 5432
; LENGTH: 627
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae

```

US-10-467-657-5432

Query Match 2.6%; Score 104; DB 6; Length 627;
 Best Local Similarity 19.7%; Pred. No. 1.2;
 Matches 124; Conservative 87; Mismatches 208; Indels 210; Gaps 33

| | | | |
|----|-----|--|-----|
| Qy | 241 | KLQEGGASSLAPDLRAQIDLTLEITPRYYVLELGLPLGDDYAAKRLINGLSGVNIIWS | 300 |
| Db | 10 | KVIENAEQARTTPTSIAYLDG--GEI-----LVGAP-----AKR-QAVTNKNTIYA | 53 |
| Qy | 301 | VGGGGASALVGLTRFKFMNEAFIR-----MTAAEQVDLFV-----ATPSNIPAES | 346 |
| Db | 54 | -----AKRLIG-----HKPDEKGVQDIESMPFEIIKADNGDAWKAQKELSPPOISAEV | 104 |
| Qy | 347 | FEVYEVALALVAQAFIGKPKPHLLQADAKQFOOLQAQKVMAMEIPAMLYDTNNWEIDFGL | 406 |
| Db | 105 | LKKKEA-----AEAVLGEK-----VTEAVITVPAYFNDSQOATKDAGR | 144 |
| Qy | 407 | ERGL-CALLIGKVDCEKRWLGDSSEDQYRNPATVEF-----VLNSNRDDNDLDUPG | 457 |
| Db | 145 | IAGLDVKRIINEPTAAALAFMGDKGDNKDKRIAYVYDLGGGTFDISIIEIANLDGDKQPEV | 204 |
| Qy | 458 | LCKLETLWLAGVFP-PRFRDTHKKFKLG--DYDDPWLVSYLE-----RVEVVGQSP | 507 |
| Db | 205 | LATNGDTFLGGDEDFQRLIDYIDEFKKEQIDILQDVMALQRLKEAABAKAKIELSSGQQ | 264 |
| Qy | 508 | L--AAAAATWARI GAEHVKASAMALOKVPSPRYTDRNSAEPKDVQETVFSVDPVGNVVG | 564 |
| Db | 265 | TEINLPYITMDATGPKHL-----AMKITRAKTESLVED-----297 | |
| Qy | 565 | RDGEPGVFIAEAVRPSENFETNDYAIR-AGVSESSVDET-----TVEMSVADMLK | 613 |
| Db | 298 | -----LIARSIEPCT-----ALKDAGLSTGDIIDDVLVGGQSRMPKVQEAOKDFPG | 344 |
| Qy | 614 | EASVK-----ILAAGVAIGLISLFSQKFIKSSSSFORXDM-----VSSMESDVATIGSV | 663 |
| Db | 345 | KEPRKDVNPDEAVAGAAI-----QGEVLSGG-----RSDVLLLDVTPLSLGIETMGGV | 393 |
| Qy | 664 | RADDSEALPRMDARTAE--IVSKWQIKSLAFQPDHRIEMLPVLQGRMLKIWTDRAAE | 721 |
| Db | 394 | -----MTKLIQKNTTIPTKASQVFSTA--EDNQSAVTHIVLQGE-----RERASA | 436 |
| Qy | 722 | TAQLG-----LVYDVT--LKLKLV-----DSVTVSADGTRAL--- | 751 |
| Db | 437 | NKSLGQNFNLGDIAPAPRGMQIETVTFDIDANGILHVSADKDKGTGKAANITIQGSSGLSEE | 496 |
| Qy | 752 | -----VEATLEESACLSDLVHPENNA | 772 |
| Db | 497 | RIRBMVXDAANAAFEEDKKIPELVASRNOA | 525 |

RESULT 9
 US-10-467-657-70
 Sequence 70, Application US/10467657
 Publication No. US20050260581A1
 GENERAL INFORMATION:
 APPLICANT: CHIRON SpA
 APPLICANT: FONTANA Maria Rita
 APPLICANT: PIZZA Mariagrazia
 APPLICANT: MASIGNANI Vega
 APPLICANT: MONACI Elisabetta
 TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
 FILE REFERENCE:
 CURRENT APPLICATION NUMBER: US/10/467,657
 CURRENT FILING DATE: 2003-08-11
 PRIOR APPLICATION NUMBER: GB-0103424.8
 PRIOR FILING DATE: 2001-02-12
 NUMBER OF SEQ ID NOS: 9218
 SOFTWARE: SeqWin99, version 1.04
 SEQ ID NO 70
 LENGTH: 715
 TYPE: PRT
 ORGANISM: Neisseria gonorrhoeae

US-10-467-657-70

| Query Match | 2.5%; | Score | 103.5; | DB | 6; | Length | 715; |
|-----------------------|--------|---|----------|------------|------|--------|------|
| Best Local Similarity | 21.6%; | Pred. | No. 1.6; | | | | |
| Matches | 126; | Conservative | 80; | Mismatches | 189; | Indels | 189; |
| Gaps | 36; | | | | | | |
| Qy | 333 | DLFVA--TPSNIPASFEFYEVALALVAQAFT-----GKPHLLQDADKQFQQLQQAQV | 384 | | | | |
| Db | 147 | DLFVIDAPCIYPRKTLGGGLYALEVMQATTKRVLDTGRMP-----LPDPTAYQQLLHG-M | 202 | | | | |
| Qy | 385 | NAMEIPA--MLYDTRNN-----WEIDFGLERGLCAL-----LIG | 416 | | | | |
| Db | 203 | AAVDYTADELIIYRSNNRSKYVGYGSPVEQIIMTVNIALKQVHALEYTYTAGSPDALVG | 262 | | | | |
| Qy | 417 | -----KUDECR-----MW-LGLDSEDSQYRNPAIVFVLENSRDNDDLPGLCKLLETWL | 466 | | | | |
| Db | 263 | VPETWSADDIRRFOEYMDLLSGTGAQRK--MRFV----- | 296 | | | | |
| Qy | 467 | AGVVFPRFDTKDKFKLGDYDDPMVLSTYLERVEV-----VQGSPLAAAA--TWARIGAE | 520 | | | | |
| Db | 297 | PGELSRNFRETKQPLK--DYDE-----WLARVVCFAPFSVEPTFPFAQVNRSAVETSR | 349 | | | | |
| Qy | 521 | HVKASAMQALQKVPF-----SRYTD-----RNSAEPKDVQFT-----VFS | 555 | | | | |
| Db | 350 | QSLSDGMSGLKNWVKALITDVLARYMDMAAYEFVWKGEESLNPKQEAIIYIYKNAGILT | 409 | | | | |
| Qy | 556 | VDPGNNVGRD-----GEPGVFIAEAVRPPSPNFENDYAIRAGYSESSVDETTVMVSAD | 610 | | | | |
| Db | 410 | ADETRAEELGKEPLPGQGOPEPKQDGRKPEE--PPNQAEKLGKSESMPSE---DESA-- | 463 | | | | |
| Qy | 611 | MLKEASVKILAAAGVAIGLISLFOSKIFLKSSSSSFQKMDVSMESDVATIG-----S | 662 | | | | |
| Db | 464 | LIEAYLLTRIDGLAEQIAALIE-----GAAVDMQAGDLAAELSRAGVANGLDFGDWS | 517 | | | | |
| Qy | 663 | VRADDSEALPRMDARTAEINVSQKQIKSLAFGPDHRIEMTPEVLDGRMLKI-----WT | 716 | | | | |
| Db | 518 | GLSDWVEPIIR-----RVAED-----GAVAL-----LRVMEPEAAGMVTNIRSAVKWA | 562 | | | | |
| Qy | 717 | -DRAAETAQL-----GLAVDYTLKLSDVSTVSADGTRALVEATLEESACLSDLV----- | 766 | | | | |
| Db | 563 | HERAEAEVWGKRACGLVRN-----PAAEQWIT-----EGTREMIRAQVAEMRNGDSVQELA | 615 | | | | |
| Qy | 767 | -----HPENNA-----TDVRTVTYTRYEFVMSKGWKITEGSVLAS | 801 | | | | |
| Db | 616 | GRLKESHAFGNARARTIARTETAMADGNGNLIGW-----EGTGLVA | 656 | | | | |

| Matches | 126; | Conservative | 80; | Mismatches | 189; | Indels | 189; | Gaps | 36; |
|---------|------|--|-----|------------|------|--------|------|------|-----|
| Qy | 333 | DLEVA--TPSNIPAESEFVEYVALALVAQAFI-----GKPHLLQADQKQFOOLOQAKV | 384 | | | | | | |
| Db | 147 | DLFVIDAPCIYPRKTLGGGLYALEVWDGATIKRVLDNTGRMP--LPPDTAYQIILHG-M | 202 | | | | | | |
| Qy | 385 | NAMEIFA--MLYDTRNN-----WEIDFGLERGLCAL-----LIG | 416 | | | | | | |
| Db | 203 | AAVDYTADELIYRSRNRYSKYVGYGSPVEQIIMTNVIALKQVHALEYTAGSVPPDALVG | 262 | | | | | | |
| Qy | 417 | -----KVDECR---NW-LGLDSEDSQYNPAIVEFVLENSNRDNDLPLGLCKLLETWL | 466 | | | | | | |
| Db | 263 | VPTWSADDIRRFQETWDLLLSGETAQRK-----MRFV----- | 296 | | | | | | |
| Qy | 467 | AGVVPFRDTKDKFKLGDYDDPMVLSYLERVEV-----VOGSPLAAAA--TMARIGAE | 520 | | | | | | |
| Db | 297 | PGELSRNFRETKQPLK--DVYDE-----WLARVVCFAFSVEPTPFVAQVNRSAVETSRE | 349 | | | | | | |
| Qy | 521 | HVKASAMQALQKVP-----SRYTD-----RNSAEKDVQET-----VFS | 555 | | | | | | |
| Db | 350 | QSLSDGMSGSLKNVWKALIDVLARYNDMAAYEFVWKGEESLNPKEQAEIYAIYKNAGILT | 409 | | | | | | |
| Qy | 556 | VDVPGNNVGRD-----GEPGVFAEAVRPSNPETNDYAIRAGVSSSVDETTVENSVD | 610 | | | | | | |
| Db | 410 | ADIEIRAEIKGKPLPGQGEPEPKQGRKEEE--PPNQGAELKGSSEPMSE---DESAA- | 463 | | | | | | |
| Qy | 611 | MLKEASVKILAAAGVAIGLISLFQKYSKSSSFQKDMVSMSESDVATIG-----S | 662 | | | | | | |
| Db | 464 | LIEAYLLTRIDGLAEQIAALIE-----GAAVDWQAGDLAAELSRAGVYVANGLDPGDWS | 517 | | | | | | |
| Qy | 563 | VRADDSALPRMDAPTAENIVSKWKIKSLAGCPDHRIEMLPRVLDGRMLKI-----WT | 716 | | | | | | |
| Db | 518 | GLSDVVEPIIR---RVAED-----GAVAAL-----LRVMPEPAAQGMVNTIRSAVKWA | 562 | | | | | | |
| Qy | 717 | -DRAAEATAOL-----GLVVDYTLKLKLSVDSVTVSADGTRALVEATEESACLSDLV- | 766 | | | | | | |
| Db | 563 | HERAAEWGMRAGGGLVRN-----PAAEQWIT---EGTREMIPAQVAEAEWNGDSVQELA | 615 | | | | | | |
| Qy | 767 | -----HPENNA---TDVRYTYTRYEVFVWSKGWKITEGSLVAS | 801 | | | | | | |
| Db | 616 | GRUKESHAFGNARPARTIARTETAMADGMNLIWG---EGTGLVA | 656 | | | | | | |

Query Match 2.5%; Score 103.5; DB 6; Length 715;
Best Local Similarity 21.6%; Pred. No. 1.6;

Db 712 KALREHAESIKKRLAESIGGLAERVDVAASVETAPDKAQMLLSQRVGDFDGTGKITLV 771
QY 190 GEALLKRLPKSPKQDVVVLWALAFLOVSRDAMALDPDPFITGYEFVEEALK-----241
Db 772 AENLTPEP-----AVMAAHHELGHGRFAAD-----GFAKYREELERADGNGLI 814
QY 242 -----LLQE--EGASSLAPDLR-AQIDETLEETPRVYVLELLGLPLGDDVAARKLNGLSG 293
Db 815 RRIADAVQEGREGTGDAAASVRPAAVEAVAL-----YAAQRTGGWAG 858
QY 294 VRNLTW-SVGGG-----GASALVGGGLTR---EKFMNEAFLRMTAAEQVDLFFVATPSNIP 343
Db 859 IENRYGVKGNLGRGIAGVLARIGALLRVLQRLAGKAGGAMSDADVAFMLADLHGNGVE 918
QY 344 AESFEVY--EVALALVAQAFTGKPKHLLQADKQFOOLQQAQKVMAMEIP--AMLYDTRNN 399
Db 919 GARDAPMGHNRHRAVMPFAEDGAERSSSEKLEKLRRAETIRISGREVPPEGNLRREYKRN 978
QY 400 WEIDFGLERGLCALLIGKVDECRMWLGL-----DSEDSQY-RNPAIVEFVLENS--447
Db 979 -ALEYG--KSLRGPVYNKDTGREISLGRSGITETILRHDYKDAEHLQSIATAIPQIIENAVY 1035
QY 448 -----NRD--DNDDLPGCLKLLETWLAGVVF-----PRFRDTKDKKFK 483
Db 1036 IDTLPNEDLAKNGDIQG---YEVYVSGLVNNGGADYTVRAAVAVSRNGNRYYYDHKLTKIE 1091
QY 484 LGDYDDPMVLSYLERVEVV-----QGSPL-----AAAAATWARI GAHVK 523
Db 1092 KGN-----LLSLDRVSTTGASESKSPLSGIDDKRLQLQDKDAGKGGIADPDTAVR 1145
QY 524 ----ASAMQALQVPSRYTDRNSAEPKQVETVFSVDPVGNVGRDGEVGF--IAEAV 577
Db 1146 FSRAANIEAAGRITGKKSRLNALKDRWDASKGIQLQFLGRRQIEDIYGGVLDGLKEVG 1205
QY 578 RPSENFTNDYATRAGVSESSVDETTVMESVADMLKEASVKILAAGVAIGLISLFSQKYF 637
Db 1310 KAAYALEKAQE-----KHGREILADEADMLRFLFYADSEAKRA-----LRA 1352
QY 737 SVDSVTVS---ADGTRALVEA 754
Db 1353 GADVAAESRAKTDVARMLEQA 1373

RESULT 12

US-10-467-657-6322
; Sequence 6322, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 6322
; LENGTH: 2004
; TYPE: PR1
; ORGANISM: Neisseria gonorrhoeae

US-10-467-657-6322

Query Match 2.5%; Score 103.5; DB 6; Length 2004;
Best Local Similarity 19.6%; Pred. No. 7.8;
Matches 145; Conservative 108; Mismatches 279; Indels 209; Gaps 36;

QY 144 RSREYNE-----GLDDDEEAFTVDVPMWKVPYGCALCVLQEG--GETEIVLRV 189
Db 712 KALREHAESIKKRLAESIGGLAERVDVAASVETAPDKAQMLLSQRVGDFDGTGKITLV 771
QY 190 GEALLKRLPKSPKQDVVVLWALAFLOVSRDAMALDPDPFITGYEFVEEALK-----241
Db 772 AENLTPEP-----AVMAAHHELGHGRFAAD-----GFAKYREELERADGNGLI 814
QY 242 -----LLQE--EGASSLAPDLR-AQIDETLEETPRVYVLELLGLPLGDDYAAKRLNGLSG 293
Db 815 RRIADAVQEGREGTGDAAASVRPAAVEAVAL-----YAAQRTGGWAG 858
QY 294 VRNLTW-SVGGG-----GASALVGGGLTR---EKFMNEAFLRMTAAEQVDLFFVATPSNIP 343
Db 859 IENRYGVKGNLGRGIAGVLARIGALLRVLQRLAGKAGGAMSDADVAFMLADLHGNGVE 918
QY 344 AESFEVY--EVALALVAQAFTGKPKHLLQADKQFOOLQQAQKVMAMEIP--AMLYDTRNN 399
Db 919 GARDAPMGHNRHRAVMPFAEDGAERSSSEKLEKLRRAETIRISGREVPPEGNLRREYKRN 978
QY 400 WEIDFGLERGLCALLIGKVDECRMWLGL-----DSEDSQY-RNPAIVEFVLENS--447
Db 979 -ALEYG--KSLRGPVYNKDTGREISLGRSGITETILRHDYKDAEHLQSIATAIPQIIENAVY 1035
QY 448 -----NRD--DNDDLPGCLKLLETWLAGVVF-----PRFRDTKDKKFK 483
Db 1036 IDTLPNEDLAKNGDIQG---YEVYVSGLVNNGGADYTVRAAVAVSRNGNRYYYDHKLTKIE 1091
QY 484 LGDYDDPMVLSYLERVEVV-----QGSPL-----AAAAATWARI GAHVK 523
Db 1092 KGN-----LLSLDRVSTTGASESKSPLSGIDDKRLQLQDKDAGKGGIADPDTAVR 1145
QY 524 ----ASAMQALQVPSRYTDRNSAEPKQVETVFSVDPVGNVGRDGEVGF--IAEAV 577
Db 1146 FSRAANIEAAGRITGKKSRLNALKDRWDASKGIQLQFLGRRQIEDIYGGVLDGLKEVG 1205
QY 578 RPSENFTNDYATRAGVSESSVDETTVMESVADMLKEASVKILAAGVAIGLISLFSQKYF 637
Db 1206 RLSELFAGD-----ANKAVTEADKVREWG---RLKEEDAKALA-----DLMDHATL 1249
QY 638 LK-SSSSFFQKD-----MVSSMESDVATIGS---VRADDSEALPRMDAR 677
Db 1250 AKVDADPLMRKDAQKRLDGI RTALDIADGKIEKAEAAVASAGARIADAAAYNKAAQRAAD 1309
QY 678 TAENIVSKWKIKSLAFGPDHRIEMLPEVLDGRMLKI-WTDRAAEFTAQLGLVYDTLLKL 736
Db 1310 KAAYALEKAQE-----KHGREILADEADMLRFLFYADSEAKRA-----LRA 1352
QY 737 SVDSVTVS---ADGTRALVEA 754
Db 1353 GADVAAESRAKTDVARMLEQA 1373

RESULT 13

US-11-074-176-98
; Sequence 98, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176


```

; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 98
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-11-074-176-98

```

```

Query Match      2.5%; Score 103; DB 7; Length 565;
Best Local Similarity 22.2%; Pred. No. 1.2;
Matches 112; Conservative 67; Mismatches 167; Indels 158; Gaps 30;

QY      200 KSFQDVVLVMAFLDVSADAMALDPDF--ITGYEFVBEALKLQEEGASSLAPDLRA 257
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      123 KSYQ-----MPLALVQIQYKFRDENRPRGLLRGFEV-----MDGVSFAA--TRE 168

QY      258 QIDETLEEIPRYVLELLGLPLGDVYAARKLNGLSG--VRNILWSVGGG-----AS 307
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      169 QLDQFPDDQKSAV-----KRIFKEAGVTVPVVIADSGTMGGKNSTEFQAP 213

QY      308 ALVGG---LTREKPMNEAFRLMTAAEOVDLF-----VATPS-NIPAESPEVYE 351
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      214 AAIGEDTIANEXGTYAANLEM--AKSIDTFKQBPEEAKELTKVATPACDITKILAELFD 271

QY      352 VALALVAQAFTI---GKKPHLLQDADKQFOOLOQAKVM-AMEIPAMLYDTNRNWEIDFGL 406
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      272 VPATRIVKSILYIADDDQVLIIRGDKQINEVKUGHVLDADDIHEA--NTEDLKEIT-GS 328

QY      407 ERGICALLIGKV-----DEC-----RMWLGLDSEDSQYRNP-----AIVEFV-L 444
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      329 EKG---GUGPVNADWADKIADETVKGLYNVVVVGAGETDYPQKNANLDRDFKVDFAFI 384

QY      445 ENSNRDNDNLLPGL-CKLLETWLAGVVPFRPRDTKDKKFKLGDYDDPMVLVSYLER---- 499
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      385 RTANEGEFPDVLHLPLKFTTTSIEVGHI-----FKLGTYYTITMGADFLDONGKA 433

QY      500 VEVVQGSPLAAATMARIGAHEHYKASAMQALQKVPPSRYTRDNSAEKPKDQVETVPSVDPV 559
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      434 QPVIMG-----YGIGVTRM-----LSAVVEOHLTDGRVAVPKRI--APPEIHIV 476

QY      560 GNNVGRDGPVGFIAEAVRP--SENFET--NDYAIRAGVS-----595
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      477 QMKMNKEDQ--TELAELKLEKFKSEKYDVLDDRRKERAGVKFADADLYGAPVRIITIGKCA 534

QY      596 -----ESSVDETTVMMSVADMLK 613
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      535 DGIIVEVKRPTDREKAVENSIDELDK 558
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 14
US-10-858-730-27
Sequence 27, Application US/10858730
Publication No. US2005025568A1
GENERAL INFORMATION:
APPLICANT: Bailey, Richard B.
APPLICANT: Blomquist, Paul
APPLICANT: Doten, Reed
APPLICANT: Driggers, Edward M.
APPLICANT: Madden, Kevin T.
APPLICANT: O'Leary, Jessica
APPLICANT: O'Toole, George
APPLICANT: Trueheart, Joshua
APPLICANT: Walbridge, Michael J.
APPLICANT: Yorgey, Peter S.
TITLE OF INVENTION: METHODS AND COM
TITLE OF INVENTION: PRODUCTION
FILE REFERENCE: 14184-030001
CURRENT APPLICATION NUMBER: US/10/8
CURRENT FILING DATE: 2004-06-01
PRIOR APPLICATION NUMBER: US 60/475

```

; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-858-730-27

```

Query Match **2.5%**; Score 101; DB 6; Length 403;
Best local Similarity 21.3%; Pred.No. 0.99;
Matches 76; Conservative 51; Mismatches 109; Indels 120; Gaps 16

```
QY      SSSSSFATATTATLVLPSPSIDRPER-----HVPIPIDFYQLVGAQTHTFLTDGIRRA 108  
Db      :|::||| | | | | :: ||:::|||::|||::|::|::|::|::|::|::|:  
        SAKFAFADITNTVKARILETGYDSSDKGFPGCATCGVNIGI-----GAQSDDIAQGVDTA 115  
QY      FEAVR-----SKPPQGFSDDALISRRQILQAACETLNSPRSRREYN EGLLD 155  
Db      |||||   |:|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
        HEARVEAGAADPLDSOGAGDQGLMEGYA---INATPELMPLIALAHLRSR----- 163  
QY      DEEATVTIDVPMDKVGALCVLOGGBSTEIVLRVGEALLKERLPKSFKODVVVLNMAAPL 215  
Db      |:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
        LTEV----RKNGVLPYLRLPDGKTQVTI-----AYEDNVPVRLDTVVISTQHAADI 209  
QY      DVSRDAWALPPPDPFITGYEFVEEAALKLOBEGASSLAPDLRAQIDETELETTPRVYLE-- 273  
Db      |:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
        DLEK----TLDFP-----IREKVLNTVLDDL---AHETLDASTVRVLVNPT 248  
QY      --LLGPLGDYYAAKRNLINGSVRNTL-:-WSVGGGGA----- 306  
Db      :||:|||||:|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
        GKFPVGMGD-----AGUTGRKKI VDTYGWARHGGGAFSGKDPSKVDRSAAYAMRW 301  
QY      ---SALVGGJLTREKFMNEAEFLRTAAEQVDLFVAT-----PSNI PAESEFEVYE 352  
Db      |:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
        VAKNVVAAGLAERVEGVQVAY-AICKAAPVGLFVFETFGTETEDPVKIEKAIGEVPDL 356
```

RESULT 15

```
US-10-467-657-2966  
 ; Sequence 2966, Application US/10467657  
 ; Publication No. US20050260581A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CHIRON SpA  
 ; APPLICANT: FONTANA Maria Rita  
 ; APPLICANT: PIZZA Mariagrazia  
 ; APPLICANT: MASIGNANI Vega  
 ; APPLICANT: MONACI Elisabetta  
 ; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
 ; FILE REFERENCE:  
 ; CURRENT APPLICATION NUMBER: US/10/467,657  
 ; CURRENT FILING DATE: 2003-08-11  
 ; PRIOR APPLICATION NUMBER: GB-0103424.8  
 ; PRIOR FILING DATE: 2001-02-12  
 ; NUMBER OF SEQ ID NOS: 9218  
 ; SOFTWARE: SeqWin99, version 1.04  
 ; SEQ ID NO 2966  
 ; LENGTH: 711  
 ; TYPE: PRt  
 ; ORGANISM: Neisseria gonorrhoeae  
 US-10-467-657-2966
```

| | | | | |
|-----------------------|--------------|---|--|-------------|
| Query Match | 2.5% | Score 101; | DB 6; | Length 711; |
| Best Local Similarity | 19.2%; | Pred. No. 2.4; | | |
| Matches 138; | Conservative | 98; | Mismatches 237; | Indels 244; |
| Gaps 35 | | | | |
| Qy | 169 | KVPGALCVLQGGETE--- | IVLRVGEALKKERLPKSPQDVLVWNAFLDVSRDAMALD | 225 |
| | | | | |
| Db | 77 | KIPGGF-FKRGKQSEKITSLRIDPRLFPFEGFYHDIQIVAMVVSVDPEIDS--- | 132 | |
| | | | | |
| Qy | 226 | PPDPIITGVFVEEAKLLQEGASSLPDLRAQKIDETLEEITTPRVLELLGLPLGDVAA | 285 | |

```
Db 133 IPAML-----|||:-----ALVLSGVFPAGPIGA 156
Qy 286 KRLNGLSGVRNILWSVGGGASALVGLTREKFMNEAFLRMTAAE-----QVDLFVATPSN 341
Db 157 ARVGYINGV-----YVLNPTKAEIAKSQDLVVAGTS- 188
Qy 342 IPAESFEVVEVALVAQAFIGKPKPHLLQDADKQFOOLOA-----KVMAMEIPAMLYDTR 397
Db 189 -----KAVLMVESEAKILPEDVMVGAVVYGHQDMQVAINAINEFADEVNPEVMDWK 239
Qy 398 ----NNWEIDFGLERGLCALLIGKVKDECRMWGLGDSSEDOYRNPAI VEFVLENSNRDND 454
Db 240 APETNEEL-VAKVKGIAGETIKEAFKIRQKARSAKLDEAWNAKKEALITTEETDLAANE 298
Qy 455 LPGCLKLETLWLAGV-----PPRF--RDTKDKFKLGDYDDPMVL--SYLERVEVV 503
Db 299 IKGIFKELE---ADVRSQILDGQPRIDGRDTRTVR-----PLNIQTGVLP--T 343
Qy 504 QGSPAAAAATWARI GAHVKASAMQALQKVPFPRYTDNRNABPKDVOETVFSVDPVG-NN 562
Db 344 HGSALFTRGETQALAVATLGTSRDEQIIDALSGEYTDRFMLH-----YNFPPYSTGE 395
Qy 563 VGRDGEP-----GVFTAEAVRPSENFETNDYAIRAGVSESSVDDETTVEMSVAD 610
Db 396 VGRWGAFFRREIGHRLAKALLAVLPKP-EDF---SYTRKV-VSE--ITESNGSSSMAS 448
Qy 611 M-----LKEASVKILA--AGVAIGLISLFSQKYFLKSSSFORKDMVSSMESDVATIGS 662
Db 449 VCGCLSLLSAGVPLKAHVAGIANGLI-----LEGNKFAVLT 485
Qy 663 VRADDSEALPRMD---ARTAEIVSKWKQIK-----SLAFGPDHRIEML----- 703
Db 486 DILGDEHLDGMDPFKVAGTTGEGVTALQMDIKIQG1TKEIMQIALAQAKEARLHILDQKA 545
Qy 704 ----PEVLGGRMLKIWTDRAE-----TAQLGLVYDYTLKLSVDSVT 742
Db 546 AVAGQBELSAHAPRLFTMKISQDKIRDVIGKGGETIRHSITAETG-----TEIN 593
Qy 743 VSADGTRALVEATLEESACLSDLVHPENNATDVRTYTRYVEVFWSKSGWKITEGSVL 799
Db 594 IAEDEGT-ITTAATQOEAG-----DAKKRIEETAEVEV-----GKVEGTUV 635
```

Search completed: December 4, 2005, 12:42:48
Job time : 15 secs

(FILE 'HOME' ENTERED AT 14:25:11 ON 04 APR 2007)

FILE 'BIOSIS, CAPLUS, CABA, AGRICOLA' ENTERED AT 14:25:30 ON 04 APR 2007

| | |
|----|--|
| L1 | 47 S FTN2 OR ARC6 |
| L2 | 22 DUPLICATE REMOVE L1 (25 DUPLICATES REMOVED) |
| L3 | 375 S PLASTID DIVISION |
| L4 | 39 S L3 AND REVIEW |
| L5 | 31 S L4 AND ENGLISH/LA |
| L6 | 28 DUPLICATE REMOVE L5 (3 DUPLICATES REMOVED) |

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 12:20:59 ; Search time 47 Seconds
(without alignments)
1409.005 Million cell updates/sec

Title: US-10-600-070B-2

Perfect score: 4063

Sequence: 1 MEALSHVGIGLSPFQLRLP.....YEVFWSKSGWKITEGSLAS 801

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5_COMB.pep.*
2: /cgn2_6/prodata/1/iaa/6_COMB.pep.*
3: /cgn2_6/prodata/1/iaa/H_COMB.pep.*
4: /cgn2_6/prodata/1/iaa/PCUTUS_COMB.pep.*
5: /cgn2_6/prodata/1/iaa/RE_COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfilese1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 145.5 | 3.6 | 6095 | 2 | US-09-144-085-2 |
| 2 | 143 | 3.5 | 2089 | 1 | US-08-418-893D-23 |
| 3 | 143 | 3.5 | 2089 | 1 | US-08-418-893D-24 |
| 4 | 139.5 | 3.4 | 954 | 2 | US-09-057-969-2 |
| 5 | 131.5 | 3.2 | 1194 | 1 | US-08-680-326-35 |
| 6 | 128.5 | 3.2 | 876 | 1 | US-08-436-664-20 |
| 7 | 128.5 | 3.2 | 876 | 1 | US-08-436-664-32 |
| 8 | 128.5 | 3.2 | 876 | 1 | US-08-436-664-34 |
| 9 | 128.5 | 3.2 | 876 | 2 | US-09-135-642-32 |
| 10 | 128.5 | 3.2 | 876 | 2 | US-09-135-642-32 |
| 11 | 128.5 | 3.2 | 876 | 2 | US-09-135-642-34 |
| 12 | 128.5 | 3.2 | 876 | 2 | US-08-394-232A-20 |
| 13 | 128.5 | 3.2 | 876 | 2 | US-08-394-232A-32 |
| 14 | 128.5 | 3.2 | 876 | 2 | US-08-394-232A-34 |
| 15 | 128.5 | 3.2 | 876 | 4 | PCT-US95-04080-20 |
| 16 | 128.5 | 3.2 | 876 | 4 | PCT-US95-04080-32 |
| 17 | 128.5 | 3.2 | 876 | 4 | PCT-US95-04080-34 |
| 18 | 126.5 | 3.1 | 2482 | 2 | US-09-252-991A-16967 |
| 19 | 125.5 | 3.1 | 5087 | 1 | US-09-144-085-1 |
| 20 | 122.5 | 3.0 | 2756 | 1 | US-08-375-709-11 |
| 21 | 122.5 | 3.0 | 2756 | 1 | US-08-752-929-11 |
| 22 | 122.5 | 3.0 | 2756 | 2 | US-09-090-793-7 |
| 23 | 122.5 | 3.0 | 2756 | 2 | US-09-231-899-7 |
| 24 | 121.5 | 3.0 | 680 | 2 | US-09-252-991A-26639 |
| 25 | 121.5 | 3.0 | 947 | 2 | US-09-540-236-1991 |
| 26 | 121.5 | 3.0 | 1220 | 2 | US-09-540-236-3011 |
| 27 | 120 | 3.0 | 852 | 2 | US-09-585-858-19 |

| | | | | | | |
|----|-------|-----|------|---|-------------------|-------------------|
| 28 | 120 | 3.0 | 852 | 2 | US-10-270-878-19 | Sequence 19, Appl |
| 29 | 120 | 3.0 | 1262 | 2 | US-09-198-452A-97 | Sequence 97, Appl |
| 30 | 120 | 3.0 | 1266 | 2 | US-09-438-185A-83 | Sequence 83, Appl |
| 31 | 119.5 | 2.9 | 831 | 1 | US-08-073-384C-5 | Sequence 5, Appl |
| 32 | 119.5 | 2.9 | 831 | 1 | US-08-254-359A-5 | Sequence 5, Appl |
| 33 | 119.5 | 2.9 | 831 | 1 | US-08-483-043-5 | Sequence 5, Appl |
| 34 | 119.5 | 2.9 | 831 | 1 | US-08-481-238-5 | Sequence 5, Appl |
| 35 | 119.5 | 2.9 | 831 | 1 | US-08-471-066B-5 | Sequence 5, Appl |
| 36 | 119.5 | 2.9 | 831 | 1 | US-08-484-956-5 | Sequence 5, Appl |
| 37 | 119.5 | 2.9 | 831 | 1 | US-08-757-653-5 | Sequence 5, Appl |
| 38 | 119.5 | 2.9 | 831 | 1 | US-08-599-491-5 | Sequence 5, Appl |
| 39 | 119.5 | 2.9 | 831 | 1 | US-08-756-386-5 | Sequence 5, Appl |
| 40 | 119.5 | 2.9 | 831 | 1 | US-08-823-516-5 | Sequence 5, Appl |
| 41 | 119.5 | 2.9 | 831 | 2 | US-08-682-853A-5 | Sequence 5, Appl |
| 42 | 119.5 | 2.9 | 831 | 2 | US-08-759-038-5 | Sequence 5, Appl |
| 43 | 119.5 | 2.9 | 831 | 2 | US-08-758-314-5 | Sequence 5, Appl |
| 44 | 119.5 | 2.9 | 831 | 2 | US-09-350-309-5 | Sequence 5, Appl |
| 45 | 119.5 | 2.9 | 831 | 2 | US-08-520-946-5 | Sequence 5, Appl |

ALIGNMENTS

RESULT 1
US-09-144-085-2
; Sequence 2, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Betlach, Mary C.
; APPLICANT: Ashley, Gary
; APPLICANT: Julien, Bryan
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 30062-20020.20
; CURRENT APPLICATION NUMBER: US/09/144,085
; CURRENT FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010,809
; EARLIER FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 6095
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-144-085-2

| | | | | |
|-----------------------|------------------|--|------------|-------------|
| Query Match | 3.6% | Score 145.5 | DB 2 | Length 6095 |
| Best Local Similarity | 19.3% | Pred. No. 0.0036 | | |
| Matches 167 | Conservative 128 | Mismatches 349 | Indels 221 | Gaps 38 |
| QY | 70 | LVSLLPSIDRPERH-VPIPIDFYQVIGAQTHFLTDGIRAREARVSKPQCFSDALIS | 128 | |
| Db | 1642 | LALLDALSPEASLVPHLDLAQ-----LQGLSGELPALF-----RALL | 1684 | |
| QY | 129 | RROLOAQCETLSNPRSRNEGLDDEATVTDVFWDKVPGALCVLQEGGETEIVLR | 188 | |
| Db | 1685 | RPSLRKXASATRRDASALRRLSALPEARLNALVELVRGEV-AAVAGLQR----- | 1734 | |
| QY | 189 | VGBALLKERLPKQKQDVLVMAALFDVSRDAMALDPDFITGYEFVEEALKLQEEGA | 248 | |
| Db | 1735 | -GRAVAADQVLKEIGLDSLMAVALRNLTSKTETSL-PATLVFDYPTPRAIAELLKQAF | 1792 | |
| QY | 249 | SSL-APDLRAQIDETLEEITPRYVLELL-----GLPLGDDVAAKRNLGSGVRNI-- | 300 | |
| Db | 1793 | SGLQVKEARVRRRRACKDEPIATVSMACRLPGGVATPDDYWRLLAEGKDAIEGLPARWD | 1852 | |
| QY | 301 | -----VGGGASALVGLTREFKFNEA-----FLRMTAAEQVD | 333 | |
| Db | 1853 | GFVEYDPPDPAAGSKSYAREGGFVRDIDLFANDPFGISPREAQSMDDPQHRVLVLETAWEAL | 1912 | |
| QY | 334 | LFVATPSNIPAESFEVVEVALA-----LVAQAFIGKKPHLLQDADKQFOQLQ | 380 | |

```
Db 1913 RAGVPSALSGSATGYLGSMGSDYGAHVTDLKELDGYRGSAASILSG-----R 1964
QY 381 QAKVMAMBIIPAMLYTRNNWEDFGELRGICALLIGKVDECRMWLGLDSEDSOYRNPAI- 439
Db 1965 VAYALGLOQPAMTVDTACSSSL-VSLHLACTALRQECB-----LALAGGVTVMTSPALF 2018
QY 440 VEFV-LENSNRDDDLPGCLKLET-----WLAG-----VFPFRFTDKKKFKLGDYVD 489
Db 2019 VEFSLKGMRSR-----GRCKSFVQADGAGWAEGCGMLLLKRLSDAQ-----RD 2063
QY 490 DPMVLSYL-----ERVEVQ-----504
Db 2064 GBRVLGVIRGSVAVNGRSGSLTAPNGPAQRVIRQALSSCGLSPEDDIDAVEAHGTGTS 2123
QY 505 GSPLAAATMARIGAE-----HVKASAMQA--LQKVPSPRYTDRNSAE 545
Db 2124 GDPLEGALAEVFGPERSPERPLYLGSSKSNLGHQAAGVAGVIMKWL-----SMQ 2175
QY 546 PKDQVETVSPVDPVGNVGRDGEVGFIAEAVRPSNFETNDYATRAGVSSSVDETTVE 605
Db 2176 HEVLKTLHAEP-SPHIGWEGS-GLSLHQEARP---WRNCRVRRAGVSSFGISGTNAH 2230
QY 606 MSVADMLKEA---SVKILAAAGVAIGLISLFSQYFLKSSSSQFQKDMVSSME---SDVAT 659
Db 2231 IILEAPAEARREPVEAEAPALLPLVLSGRDEAAVNAQAQWAKWLEEHGEVGSVDVVR 2290
QY 660 IGSVRADSEALPRMDARTAEINIVSKWQIKSLAFG-PDRIEMLPVLDGRMLKIWTD 718
Db 2291 TAAHRTHEFESASVLAASAAGV---EGLRALSGRPDAAVVSGTAKGGKLVLFQO 2347
QY 719 AAEATQAG--LVYDTLLKLSVDSVTSAD-----GTRALVEATL--ESSACL--SDLVH 767
Db 2348 GSQRLGMKRLYEVYVVFRAAFDEVCALDAYLDRLREVVFVAAAGSEBGCALLETEYQ 2407
QY 768 PENNATDVRTYTRVEVFWKSGWK 792
Db 2408 PGLFALEVALYRQ-----WESWGLK 2427

RESULT 2
US-08-418-893D-23
; Sequence 23, Application US/08418893D
; Patent No. 5559220
; GENERAL INFORMATION:
; APPLICANT: ROESSLER, PAUL G
; APPLICANT: OHROGGE, JOHN B
; TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
; TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTILLA CRYPTICA
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESS: NATIONAL RENEWABLE ENERGY LABORATORY
; STREET: 1617 Cole Blvd.
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,893D
; FILING DATE: April 7, 1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/104,938
; FILING DATE: September 14, 1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: O'CONNOR, EDNA
; REGISTRATION NUMBER: 29,252
```

```
; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-231-1000
; TELEFAX: 303-231-1098
; TELEX:
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2089 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-418-893D-23

Query Match 3.5%; Score 143; DB 1; Length 2089;
Best Local Similarity 20.0%; Pred. No. 0.00096;
Matches 169; Conservative 113; Mismatches 289; Indels 272; Gaps 44;

QY 73 LPPSIDRPERHVPIDFYQVLGAQTHFTDGIIRAFPEARVSKPP-----QFQSDAL 126
Db 197 LPNALDK-----LGIKFIGPTGPMVSLGDKIAANILAQTAKVPSIPWSGSGFGPDDB 250
QY 127 ISRRQLQACETLSNPSRREYNEGLD--DEEATVITDPWDKVPALCVLEGG--- 181
Db 251 LQ-----ADLTEGTIPMEIFNKLVTSADEAVIVANKIGHEN--GIMIKASEGGGK 301
QY 182 -----ETEIVLR-----VGEALLKERLPKSKQDVVLVMAFLDVSRDAMALDP 226
Db 302 GTRFVNEADLRNAFVQVNEVIGSIFLMQLCKNARHIEVQIVG-----DOHGNVALNG 357
QY 227 PDFITGYEFVEEALKLQEBEGASSLAPDLRAQIDETLEEITPRYVLELLGLPLGDDYAAK 286
Db 358 RDCSTQRRF-----QKIFEEGPPSIVP-----KETFHEM-----ELAAQ 391
QY 287 RLNGLSGVNLSWVGCGGASALVGLITREKFNNEAFLRMTAAEQVDLFVA---TPSNIP 343
Db 392 RL-----TQNGYQ--GAGTVEYLYNAADNKKF---FLEINPRLQVEHPVTEGITGANLP 441
QY 344 AES-----PEVYEVALALVAQAFIGKPKHLLQADAKQFOOLOQAQKMAEIPA--- 391
Db 442 ATOLQVAMGIPUNIPDIRRLYGREDAYGTD--IDFLQERYREL--DSHVTAARITANP 498
QY 392 -----MLYDTRNNW--EIDFGLERGLCALLIGKVDCEMRMLGDSDSQV--- 434
Db 499 DEGFKPTSGSIERIKFQSTPNVWGYFVGANGI-----HEFADSQEGH 542
QY 435 ---RNP-----AIVEFVLENSNRDD--NDDLPGCLKLET-----WLAGVVP 472
Db 543 LPAKGNPNEQARKALVLALKEMEVRGDIRNSVEYLVKLETEAFKNTIDTSLWDGII-- 600
QY 473 RFRDTHDKKFKLGDYDDPMVLSYLERVEVQCSPLAAATMARIGAHHVKASAMQA--- 529
Db 601 -----KEKSVKV-----EMPSHL-----VVVGAAPFAFEHVKVAETEEVKESPRKQVS 644
QY 530 -----LQKVPSPRYTDRNSAEPKQVQET----- 552
Db 645 TAGIPGINSFNIEVAYLDTKYPFHVERISPDVTRPTLDGNTIDVEVTQTAEGALLATFG 704
QY 553 -----VFSVD--PVGNNVGRDGE-----PGVFIAEAVRPS-----ENFTNDYAIRAGVSES 597
Db 705 ETHRIFGMDEPLGLRLSLDGATVLMPTIFDPSELRTDVTGKVVRVYLDQNGATVEAG--- 760
QY 598 SVDETTVMSVADMLKEASVKILAAAGVAIGLISLFSQYFLKSSSSQFQKDMVSSME--SD 656
Db 761 ---QPYVEVEAMKMI--MPIKATESG-----KITHNLSAGSVISAGDGLASLELD 806
QY 657 VATIGSV-----RADDSEALPRMDARTA--ENIVSKWQIKSLAFGPDRIEMLPVLDGR 710
Db 807 PSRVKKIETFSKGLDIMESKVDLEPQKAVNVNLS-----GFNLD-----PEAVAQ 852
QY 711 MLKIWTDRAAETAQLGLVYD--YTLKLSVDSVTSVADGTRALVEATLEESACLSDLVHPE 769
```

Db 853 AIDSATDSSAAADLLVQLDFEYRVESQFDGV-IADDVVRTLKANTE-----TLDVWSE 907
QY 770 NNA 772
Db 908 NLA 910

RESULT 3

US-08-418-893D-24
; Sequence 24, Application US/08418893D
; Patent No. 5559220
; GENERAL INFORMATION:
; APPLICANT: ROESSLER, PAUL G
; APPLICANT: OHLROGGE, JOHN B
; TITLE OF INVENTION: GENE THAT ENCODES ACETYL-CORNZYME A
; TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTELLA CRYPTICA
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
; STREET: 1617 Cole Blvd.
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,893D
; FILING DATE: April 7, 1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/104,938
; FILING DATE: September 14, 1993

CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: O'CONNOR, EDNA
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-231-1000
; TELEFAX: 303-231-1098
; TELEX:
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2089 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal

US-08-418-893D-24

Query Match

Best Local Similarity 3.5%; Score 143; DB 1; Length 2089;

Matches 169; Conservative 113; Mismatches 289; Indels 272; Gaps 44;

QY 73 LPSPIDRPRHRVPIPIIDFYQVLGAQTHFTDGIIRAFARVSKPP-----QFGFSDDAL 126
Db 197 LPNALDK-----LGKFGPTGTPVMSVLGDKIAANILAQTAQVPSIPWSGSGFGGPDGP 250
QY 127 ISRRQILQACSTLGNPRSRREYNEGLD--DEEATVITDVPWDKVPGLCVLQEGG--- 181
Db 251 LQ-----ADLTEGTIPMEIFNKGVLTSDEAVIVANKIGWEN--GIMIRASEGGGK 301
QY 182 -----ETEVLV-----VGEALKERLPKQKQDVVLVMAFLDVRDAMALDP 226
Db 302 GIRFVDNEADLRNAPVQVSNEVIGSPFIPLMQLCKNARHIEVQIVG-----DOHGNVALNG 357

QY 227 PDFITGVFEVEALKLQEGGASSLAPDLRAQIDETLEETIPRYVLELLGLPLGDDYAK 286
Db 358 RDCSTORRF-----QKIFEGPPSIVP-----KETFHEM-----ELAAQ 391
QY 287 RLNLGSLGVRNILWSVGGGASALVGLTREKPMNEAFLRMTAAEQVDLFVA---TPSNIP 343
Db 392 RL-----TONIGYQ--GAGTVEYLYNAADNKF--FLELNPRLQVEHPVTGIGTANLP 441
QY 344 AES-----FEVYEVALVAQAFIGKPKPHLLQDADKQFQQLQQAQVMAEIPA--- 391
Db 442 ATQLQVAMGIPLFNIPDIRRLYGREDAVGTDP--IDFLQERYREL--DSHVIAARITAENP 498
QY 392 -----MLYDTRNNW--EIDFGLERGLCALLIGKVDCEMRMLGLDSEDSQY--- 434
Db 499 DEGFKPTSGSTERIKFQSTPNVMGYFVSGANGGI-----HEFADSQFQGH 542
QY 435 ---RNP-----AIVFVLENSNRDD--NDLPLGLCKLET-----WLAGVVFP 472
Db 543 LFAKGNRQARKALVLAKEMEVGRGDIRNSVEYLVKLLETEAFKNTIDTSWLDGII-- 600
QY 473 RFRDTKDKFKLGDYDDPMVLSYLERVEVVGQSPLAATAATMARIGAEHVKASAMQA--- 529
Db 601 -----KEKSVKV-----EMPSHL-----VVGAAPKAFKFEHVKAATEEVKESPRKGQVS 644
QY 530 -----LQKVPFRYTRNSAEPKDVQET----- 552
Db 645 TAGIPGINSFNIEVAYLDTKYPFHEVERISPDVYRFTLDGNTIDVEVTQAEGLIATFGG 704
QY 553 ---VFSVD--PVGNVVRDGE---PGVFIAEVRPS-----ENFENDVAIRAGVSES 597
Db 705 ETHRIFGMBPLGLRLSLDGAIVLMTIFDPSSELRDVTGKVRVYLQDNGATVEAG---- 760
QY 598 SVDETTVMESVADMLKEASVKILAAQVAIGLISLFSQKYLKSSSSSFKDKMVSME--SD 656
Db 761 ---QPYVEVEAMKI--MPIKATESG-----KITHNLASGVSISAGDILLASLELKD 806
QY 657 VATIGSV-----RADDSEALPRMDART--ENIVSKWQKIKSLAFPGPHRIEMLPEVLDR 710
Db 807 PSRVKKTETFGKLDIMESKVDLEPOKAVMNVLS-----GFNLD-----PEAVAQQ 852
QY 711 MLKIWTDRAETAQLGLVYD--YTLKLSVSVTVVSADGTRALVEATELESACLSDLVHPE 769
Db 853 AIDSATDSSAAADLLVQLDFEYRVESQFDGV-IADDVVRTLKANTE-----TLDVWSE 907
QY 770 NNA 772
Db 908 NLA 910

RESULT 4

US-09-057-969-2
; Sequence 2, Application US/09057969
; Patent No. 6013451
; GENERAL INFORMATION:
; APPLICANT: WONG, VICTOR THI WONG
; APPLICANT: PHANG, SENG MENG
; APPLICANT: TAN, TIEN CHYE
; TITLE OF INVENTION: BACILLUS STEAROTHERMAPHILUS DNA
; TITLE OF INVENTION: POLYMERASE I (KLENOW) CLONES INCLUDING THOSE REDUCED TO
; TITLE OF INVENTION: 3'-TO-5' EXONUCLEASE ACTIVITY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
; STREET: 745 FIFTH AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10151
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

Db 662 LAHIAE---DDNLIEAFRRGLDHTKTAMDIFHVSEEDVTANWRRQAKVNFIVIGISD 718
QY 635 KYFLKSSSSFORQDMVSSMESDVATIGSVRADDSEALPRMDARTAEINIVSKWKIKSLAF 694
Db 719 -YGLAQNLTIRKEAAEFIERFASPPGVK-----QYMDNIVQE-AKOKGYVT 764
QY 695 GPDHRIEMLPEVLDGRM-LKIWTDRAA-ETAQLGLVYDYLKLSVD-SVTVSADGTRA- 750
Db 765 TLLHRRYLPDITSRNFNVSFAERTAMNTPIQGSAAD-IKKAMIDLVSRLREERLQAR 823
QY 751 -----LVEATLEESACLSDLVHPE 769
Db 824 LLLQVHDELILEAPKEIEIRLCRLV-PE 850

RESULT 7
US-08-436-664-32
; Sequence 32, Application US/08436664
; Patent No. 5874282
; GENERAL INFORMATION:
; APPLICANT: RIGGS, MICHAEL G.
; APPLICANT: SIVARAM, MATHOOR
; APPLICANT: TUDOR, STARLA D.
; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS
; TITLE OF INVENTION: STEAROTHERMOPHILUS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gen-Probe Incorporated
; STREET: 9880 Campus Point Drive
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,664
FILING DATE: 08-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: 08/394,232
FILING DATE:
APPLICATION NUMBER: 08/307,410
FILING DATE: 16-SEP-1994
APPLICATION NUMBER: 08/222,612
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fisher, Carlos A.
REGISTRATION NUMBER: 36,510
REFERENCE/DOCKET NUMBER: GP94003.CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-2807
TELEFAX: 619-546-7929
TELEX:
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 876 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
US-08-436-664-32

Query Match 3.2%; Score 128.5; DB 1; Length 876;
Best Local Similarity 21.0%; Pred. No. 0.0056;

Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;
QY 99 HFUTDGIRAFARVAKPQFGSDDALISRRLOIQAACETLNSRPSRREY-NEGLLDDE 157
Db 107 HVEADDIIIGTMAARE---GFAVKVISGDRDLTQ-----LASPQVTVEITKKGITDIE 158
QY 158 EATVITDV-----PWKVPQ-----ALCVLOEGGETEIVLRVG 190
Db 159 SYTPETVWEKYGTPEQIVDLKGLMGDKSDNIPGVGIGBKTAVKLLKQFGQTVENVLASI 218
QY 191 EALLKERLPKSFQ--DVLVWALAFLDVSRDA---MALDPPDPFITGVFVEEAKLLQE 245
Db 219 DEIKGEKLENLQRYRDLAL-LSKQLAAICRDPVELTLD--DIVYGEDREKVVALLQE 275
QY 246 EGASSIAPDLRAQIDE-----TLEEITPRYVLELGL---PLGDDYAAAKRLNG- 290
Db 276 LGFQSFLLDKMAVQTDGEKPLAGMDFADADSVTDEMLADKAALVVEVVDNHYHAPIVG 335
QY 291 -----LSGVRNLSVGGGASALVUGLITREKPMNEAFLMTAAE----- 330
Db 336 ALANERGRFRLPETALADPKFLAW-----LGDETKKTKTFDSKRAAVALKWKGI 385
QY 331 ----QVDLFVATPSNIPAES-----PEVVEVALALVAQAFIGK-----KPHLLQ 370
Db 386 ELRGVFDLLIAAYLLDPAQAGDVAACVAKHQVEAVRSDEAVYGVKAKRTVPDEPTLAE 445
QY 371 DADKQFQQLQQAQKAMAEIPAMLYDTRNNWE-IDFGLERGLCALLIG-----KVDECRM 423
Db 446 HLAR-----KAAAIWALEEPLMDELRRNEQDRULLTELEQPLAGILANWEEFTGVKVDTKRL 500
QY 424 -WLGLD-SEDSQYRNPAIVFVLENSNRDNDLPLGLCKLLETWLAGVVPFRFRTDKK 481
Db 501 EQMGAELEQLQ-----AVERRIYELAGQEFNINSP---KQLGT---VLEFDKLQLPVLKK 549
QY 482 FKLGDDYDDPMVLSYL-ERVEVVGQSPAAAAATWARGAEHVKASAMAQALQKVPFSRVD 540
Db 550 TKTG-YSTSADVLEKJAPHHEIVE--HILHYRQLGKLQSTYIE-GLLKVHPVTKVHTM 605
QY 541 RNSAEPRDVOET--VFSVDPVGNVGRDGPVFIABAVRPE-----NFTNDYAI 590
Db 606 FNOA-----LTQTGLSSVEPNLQNIPIRLEGRKIRQAFVSPEDWLIFAADYQIELRV 661
QY 591 RAGVSSSVDETTVE-----MSVADMLKE---ASVKILAAGVAIGLISLSQ 634
Db 662 LAHIAE---DDNLIEAFRRGLDHTKTAMDIFHVSEEDVTANWRRQAKVNFIVIGISD 718
QY 635 KYFLKSSSSFORQDMVSSMESDVATIGSVRADDSEALPRMDARTAEINIVSKWKIKSLAF 694
Db 719 -YGLAQNLTIRKEAAEFIERFASPPGVK-----QYMDNIVQE-AKOKGYVT 764
QY 695 GPDHRIEMLPEVLDGRM-LKIWTDRAA-ETAQLGLVYDYLKLSVD-SVTVSADGTRA- 750
Db 765 TLLHRRYLPDITSRNFNVSFAERTAMNTPIQGSAAD-IKKAMIDLVSRLREERLQAR 823
QY 751 -----LVEATLEESACLSDLVHPE 769
Db 824 LLLQVHDELILEAPKEIEIRLCRLV-PE 850

RESULT 8
US-08-436-664-34
; Sequence 34, Application US/08436664
; Patent No. 5874282
; GENERAL INFORMATION:
; APPLICANT: RIGGS, MICHAEL G.
; APPLICANT: SIVARAM, MATHOOR
; APPLICANT: TUDOR, STARLA D.
; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS
; TITLE OF INVENTION: STEAROTHERMOPHILUS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gen-Probe Incorporated
; STREET: 9880 Campus Point Drive

CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,664
FILING DATE: 08-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/394,232
FILING DATE:
APPLICATION NUMBER: 08/307,410
FILING DATE: 16-SEP-1994
APPLICATION NUMBER: 08/222,612
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fisher, Carlos A
REGISTRATION NUMBER: 36,510
REFERENCE/DOCKET NUMBER: GP94003.CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-2807
TELEFAX: 619-546-7929
TELEX:
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 876 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-436-664-34
Query Match 3.2%; Score 128.5; DB 1; Length 876;
Best Local Similarity 21.0%; Pred. No. 0.0056;
Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;
QY 99 HFLTDCIRRAFEARVSKPQFGSDALISRRQILQAACETLSNPRSRREY-NEGLLDDE 157
DB 107 HYEADDIIGTMAARE--GFAVKVISGDRDLTQ-----LASQVTVETIKKGITDIE 158
QY 158 EATVITDV-----PMDKVFG-----ALCVLQEGGETEIVLRVG 190
DB 159 SVTPETVVEKYGLTPBQIVDLKGLMGDKSDNIPGVPGIGETAKVLLKQFGTVENVLASI 218
QY 191 EALLKERLPKSKQ--DVLVWALAFDVSRA---MALDPPFTTGYFVEBALKLLQE 245
DB 219 DEIKGEKLENLRQYRDAL-LSKQLAAICRDAPELVLTLD--DIVYKGEDREKVALFOE 275
QY 246 ECASSIAPDLRAQIDB-----TLEETTPRVVLELGI---PLGDDYAAKRLNG- 290
DB 276 LGFQSLDKMAVQTDGEKPLAGMDPAIDSVSDTDEMLADKAALVVEVGDVNHHPVIGI 335
QY 291 -----LSGVRNLSVGGGASALVGGTLTRKFMNEAFLRMTAAE---- 330
DB 336 ALANERGRFFLPETALADPKFLAW-----LGDETKKTKMTDSKRAVALKWKGI 385
QY 331 -----QVDLFVATPNIPAES-----FEVEVALALVAQAFIGK-----KPHLLQ 370
DB 386 ELRGVVFLLIAAYLLDPAQAAGDVAAVAKMKQYEAVRSDEAVYKGAKRTVPDEPTLAE 445
QY 371 DADKQFQLOQAQKVMAMEIPAMLYDTRNNWE-IDFGLERGLCALLIG-----KVDECRM 423
DB 446 HJAR-----KAAAIWALEEPLMLRNEODRLLTLEQLPGLAGILANMEFTGVKVDTKKL 500

424 -WLGLD-SEDSOYRNPALVEFVLENSNRDDNDLPLCLCKLLETWLAGVVFPRPRDTKDKK 481
501 EOMGAELTEQLQ-----AVERRIYELAQOEFNINSP---KQLGT-----VLFDKLQLPVLKK 549
482 FKLGDYDDPMVLSYL-ERVEVVOGSPLAAAAATMARIAGAEHVKASAMQALQKVPFSRYTD 540
550 TKTG-YSTSADVLEKLAHPHIVE--HILHVRQLKQLASTYIE-GLLKVVHPVTGVKHTM 605
541 RNSABPKDVQET--VFSVDPVGNVGRDGPVFIAEAVRPSB-----NFTNDYAI 590
606 FNQA-----LTQTRGLSSVEPNLQIPRIEGRKIRQAFVPSPDWLIFAADYSQIELRV 661
591 RAGVSESSVDETTVE-----MSVADMLKE---ASVKILAAQVATGLISLFSQ 634
662 LAHTAE---DDNLTEAFRRGLDIHTKTAMDIPHVSEEDVTANMRQAKAVNFGIVYGISD 718
635 KYFLKSSSSSFORKDMVSSMESDVATIGSVRADDSSEALPRMDARTAEINIVSKWKIKSLAF 694
719 -YGLAQNLNITRKEAAEFIERYPASFPQVK-----QYMDNIVQE-AKQKGYVT 764
695 GPDHRIEMPEVLDRM-LKIWTDRAA-ETAQLGLVYDYTTLLKLSVD-SVTVSADGTRA- 750
765 TLLHRRRYLPDITSRNFNVSFAERTAMNTPIQSSAAD-IKKAMIDLVSRLREERLQAR 823
751 -----LVEATLESACLSDLVHPE 769
824 ILLQVHDELILEAPKBIERLCRLV-PE 850
RESULT 9
US-09-135-642-20
Sequence 20, Application US/09135642
Patent No. 6066483
GENERAL INFORMATION:
APPLICANT: RIGGS, MICHAEL G.
APPLICANT: SIVARAM, MATHOOR
APPLICANT: TUDOR, STARLA D.
TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS
TITLE OF INVENTION: STEAROTHERMOPHILUS
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gen-Probe Incorporated
STREET: 9880 Campus Point Drive
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/135,642
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/394,232
FILING DATE:
APPLICATION NUMBER: 08/307,410
FILING DATE: 16-SEP-1994
APPLICATION NUMBER: 08/222,612
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fisher, Carlos A
REGISTRATION NUMBER: 36,510
REFERENCE/DOCKET NUMBER: GP94003.CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-2807
TELEFAX: 619-546-7929
TELEX:
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:

```
; ; LENGTH: 876 amino acids
; ; TYPE: amino acid
; ; STRANDEDNESS: single
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: protein
; ; HYPOTHETICAL: NO
; ; ANTI-SENSE: NO
; ; FRAGMENT TYPE: internal
; ; ORIGINAL SOURCE:
; ; US-09-135-642-20

Query Match
Best Local Similarity 3.2%; Score 128.5; DB 2; Length 876;
Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;

QY 99 HFLTDCIRRAFAEAVSKPPQFGSDDALISRRLQQAACETLSNPRSRREY-NEGILLDDE 157
Db 107 HYEADDILGTMAAAREE---GFAVKVISGDRDLTQ-----LASPQVTVEITKKGITDIE 158
QY 158 EATVITDV-----PMDKVPQ-----ALCVLQEGGETEIVLRVG 190
Db 159 SYTPETVVEKYGLTPQIVDLKGLMGDKSDNIPGVPGIGEKTAVKLLKQFVTENVLASI 218
QY 191 EALLKERLPKSKQ--DVLVWALAFLDVSRDA--MALDPPDFITGYEFVEEALKLQOE 245
Db 219 DEIKGEKLENKQYRDLAL-LSKQLAAICRDAPVELTLD--DIVYKGEDREKVVAFQOE 275
QY 246 EGASSIAPDLRAQIDE-----TLEETPRYVLELLGL---PLGDDYAAKRLNG- 290
Db 276 LGFQSFLDKMAVQTDGEXPEKPLAGMDFAIADSVTDEMLADKAALVVEVGDNYHHPIVGI 335
QY 291 -----LSGVNRLMSVGGGASALVGLTREFKMFNEAFLRMTAAE----- 330
Db 336 ALANERGRFELPETALADPKFLAW-----LGDETKKTKTFDSKRAAVALKWKGI 385
QY 331 -----QVDLFVATPSNIPAES-----FEVYEVALALVAQAFIGK-----KPHILQ 370
Db 386 ELRGVVFDDLAAAYLDPAQAAGDVAVAAKMHQYEAVRSDAEAYVYGGAKRTVPDEPTLAE 445
QY 371 DADKQFQQLQQAQVMAWAMEIPAMLYDTRNNWE-IDFGLERGLCALLIG-----KVDECRM 423
Db 446 HILAR-----KAAAIWALEPFLMDELERNQODRLLTLEBQPLAGILANMEFTGVKVTKRL 500
QY 424 -WLGLD-SEDSQYRNPAIVEFVLENSNRDNDLPGCLKLLETWLAGVVPFRPRDKDK 481
Db 501 EQMGAEITQLOQ---AVERRIYELAGQEFNINSP---KQLGT---VLFDKLQLPVLKK 549
QY 482 FKLGDYDDPMVLSYL-ERVEVVOGSPLAAATMARIGAEHVKASAKQALQKVPSPRYTD 540
Db 550 TKTG-YSTSADVLEKLAHPHEIVE--HILHYQLGKLOSTYIE-GLLKVVHPVTVGKVHTM 605
QY 541 RNSAEPKDYQET--VFSVDVPVGNVGRDGEPPGVFIAEAVRPSB-----NPTNDYAI 590
Db 606 FNAQ-----ITQGRLSVSEPNLQNIPIRLEEGKIRQAFVPSPEPDWLIIFAADYSQELRV 661
QY 591 RAGVSSSVDETTVE-----MSVADMLKE---ASVKILAAGVAIGLISLFSQ 634
Db 662 LAHIAE---DDNLIEAFRRGLDHTKTAMDIFHVSEEDVTANNRRQAKAVNFGIVVGISD 718
QY 635 KYFLKSSSFQRKDMVSMESDVATTGSRVADDSREALPRMDARTANIVSKWQKIKSLAF 694
Db 719 -YCLAQNLNITRKEAAAEFTERYFASPFGVK-----QYMDNIVQOE-AKQKGYVT 764
QY 695 GPDHRIEMLEPEVLDGRM-LKIWTDRAA-ETAQGLVYVDTLLKLSVD-SVTVSADGTRA- 750
Db 765 TLLHRRYLPDITSRNPNVNSFAERTAMNTPIQGSAAD-IKKAMIDLSVRLREERLQAR 823
QY 751 -----LVEATLESACLSDLVHPE 769
Db 824 LLLQVHDELILEAPKEIERELCLV-PE 850
```

RESULT 10

```
US-09-135-642-32
; ; Sequence 32, Application US/09135642
; ; Patent No. 6066483
; ; GENERAL INFORMATION:
; ; APPLICANT: RIGGS, MICHAEL G.
; ; APPLICANT: SIVARAM, MATHOOR
; ; APPLICANT: TUDOR, STARLA D.
; ; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS
; ; TITLE OF INVENTION: STEAROTHERMOPHILUS
; ; NUMBER OF SEQUENCES: 34
; ; CORRESPONDENCE ADDRESS:
; ; ADDRESSEE: Gen-Probe Incorporated
; ; STREET: 9880 Campus Point Drive
; ; CITY: San Diego
; ; STATE: CA
; ; COUNTRY: USA
; ; ZIP: 92121
; ; COMPUTER READABLE FORM:
; ; MEDIUM TYPE: Diskette
; ; COMPUTER: IBM Compatible
; ; OPERATING SYSTEM: DOS
; ; SOFTWARE: FastSeq Version 1.5
; ; CURRENT APPLICATION DATA:
; ; APPLICATION NUMBER: US/09/135,642
; ; FILING DATE:
; ; CLASSIFICATION:
; ; PRIOR APPLICATION DATA:
; ; APPLICATION NUMBER: 08/394,232
; ; FILING DATE:
; ; APPLICATION NUMBER: 08/307,410
; ; FILING DATE: 16-SEP-1994
; ; APPLICATION NUMBER: 08/222,612
; ; FILING DATE: 16-SEP-1994
; ; ATTORNEY/AGENT INFORMATION:
; ; NAME: Fisher, Carlos A
; ; REGISTRATION NUMBER: 36,510
; ; REFERENCE/DOCKET NUMBER: GP94003.CP2
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: 619-535-2807
; ; TELEFAX: 619-546-7929
; ; TELEX:
; ; INFORMATION FOR SEQ ID NO: 32:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 876 amino acids
; ; TYPE: amino acid
; ; STRANDEDNESS: single
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: protein
; ; HYPOTHETICAL: NO
; ; ANTI-SENSE: NO
; ; FRAGMENT TYPE: internal
; ; ORIGINAL SOURCE:
; ; US-09-135-642-32

Query Match
Best Local Similarity 3.2%; Score 128.5; DB 2; Length 876;
Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;

QY 99 HFLTDCIRRAFAEAVSKPPQFGSDDALISRRLQQAACETLSNPRSRREY-NEGILLDDE 157
Db 107 HYEADDILGTMAAAREE---GFAVKVISGDRDLTQ-----LASPQVTVEITKKGITDIE 158
QY 158 EATVITDV-----PMDKVPQ-----ALCVLQEGGETEIVLRVG 190
Db 159 SYTPETVVEKYGLTPQIVDLKGLMGDKSDNIPGVPGIGEKTAVKLLKQFVTENVLASI 218
QY 191 EALLKERLPKSKQ--DVLVWALAFLDVSRDA--MALDPPDFITGYEFVEEALKLQOE 245
Db 219 DEIKGEKLENKQYRDLAL-LSKQLAAICRDAPVELTLD--DIVYKGEDREKVVAFQOE 275
QY 246 EGASSIAPDLRAQIDE-----TLEETPRYVLELLGL---PLGDDYAAKRLNG- 290
Db 276 LGFQSFLDKMAVQTDGEXPEKPLAGMDFAIADSVTDEMLADKAALVVEVGDNYHHPIVGI 335
```


Db 719 -YGLAQNINIRKRAAEFIERYFASPPGVK-----QYMDNIVQE-AKQGYVT 764
QY 695 GPDHRIEMLPEVLDRM-LKIWTDRAA-ETAQLGLVYDYLKLSVD-SVTVSADGTRA- 750
Db 765 TLHRRYLPDITSRNFNVSFAERTAMTPIQSSAAD-IKKAMIDLSVRLREERLQAR 823
QY 751 -----LVEATLEESACLSDLVHPE 769
Db 824 LLLQVHDELILEAPKEIERLCRLV-PE 850

RESULT 12
US-08-394-232A-20
; Sequence 20, Application US/08394232A
; Patent No. 6100078
; GENERAL INFORMATION:
; APPLICANT: RIGGS, MICHAEL G.
; APPLICANT: SIVARAM, MATHOOR
; APPLICANT: TUDOR, STARLA D.
; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS STEAROTHERMOPHILU
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gen-Probe Incorporated
; STREET: 9880 Campus Point Drive
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/394,232A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/307,410
; FILING DATE: 16-SEP-1994
; APPLICATION NUMBER: 08/222,612
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fisher, Carlos A
; REGISTRATION NUMBER: 36,510
; REFERENCE/DOCKET NUMBER: GP94003.CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-2807
; TELEFAX: 619-546-7929
; TELEX:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 876 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
US-08-394-232A-20

Query Match 3.2%; Score 128.5; DB 2; Length 876;
Best Local Similarity 21.0%; Pred. No. 0.0056;
Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;
QY 99 HFLTGIRAFARVSKPQFGSDALISRRLOQAACETLSNPRSRREV-NEGILLDDR 157
Db 107 HYEADDIIIGTMAARE-----GPAVKVIGSDRLTQ-----LASPQVTVETIKKGITDIE 158
QY 158 EATVITDV-----PMDKVPQ-----ALCVLQEGGETEIVLRVG 190

Db 159 SYTPETVWEKYIGTTPQIVDLKGLMGDKSNIPGVPGIGEXTAVKLLKQFGTVENVLASI 218
QY 191 EALLKERLPKSFQK--DVLVLMALAFLDVSRDA---MALDPPDFITGYEFVEEALKLQE 245
Db 219 DEIKGEKLENLQYRDLAL-LSKQLAAICRDAPELTLTLD--DIVYKGEDREKVVALLQE 275
QY 246 EGASSIAPDLRAQIDE-----TLBEITPRYVLELGL---PLGDDYAAKRLNG- 290
Db 276 LGFQSFLDKMAVQTDEGEKPLAGMDFAIADSVTDEMLADKAALVVEVVGDYNYHAPIVGI 335
QY 291 -----LSGVNIIILSVGGGASALVGLLTFREKEMNEAFLRMATAE----- 330
Db 336 ALANERGRFFLRPETALADPKFLAW-----LGDETKKTKTFMFSKRAAVALKWKGI 385
QY 331 -----QVDLFVATPSNIPAES-----FEVYEVALALVAQAFICK-----KPHLLQ 370
Db 386 ELRGVVFDLLLAAYLLDPAAGDAVAAMQHVQEAVERSDVAVYKGAKRIVPDEPTLAE 445
QY 371 DADKQFQQLQOAKVAMEIPAMLYDTNNWE-IDFGLERGLCALLIG-----KVDECRM 423
Db 446 HLAR-----KAAAIWALEEPLMDLRNEQDRLLTLEQLAGILANMEFTGVKVDTKRL 500
QY 424 -WLGLD-SEDSOYENPAIVFEVLENSNRDDNDLPGCLKLETWLAGVVFPRFRDTKDKK 481
Db 501 EQMGABLTEQLQ----AVERRIYELAQEENINSP---KQLGT----VLFDKLQLPVLKK 549
QY 482 FKLGDYDDPMVLSYL-ERVEVVQGSPLAAAAATMARIGAETHVAKASAMQALQKVPFSRYTD 540
Db 550 TKTG-YSTSADVLEKLAPHHEIVE--HILHYRLQKLGQSTYIE-GLLKVVHPVTKVHTM 605
QY 541 RNSAEPKDVQET--VFSVDPVGNVGRDGBGFVIAEAVRPE-----NFEINDYAI 590
Db 606 FNQA-----LTQTKLSSVEPNLQNIPIRLBGRKIROAFVSPEDWLIFAAADYSQIELRV 661
QY 591 RAGVSESSVDETTE-----MSVADMLKE---ASVKILAGVAIGLISLSFQ 634
Db 662 LAHIAE---DONLIEAFRRGLDIHTKTAMDIFHVSEEDVTANMRQAKAVNFVIGYISD 718
QY 635 KYFLKSSSSPQRKDMVSSMESDVATIGSVRADDSALPRMDARTAEINIVSKWQIKSLAF 694
Db 719 -YGLAQNINIRKRAAEFIERYFASPPGVK-----QYMDNIVQE-AKQGYVT 764
QY 695 GPDHRIEMLPEVLDRM-LKIWTDRAA-ETAQLGLVYDYLKLSVD-SVTVSADGTRA- 750
Db 765 TLHRRYLPDITSRNFNVSFAERTAMTPIQSSAAD-IKKAMIDLSVRLREERLQAR 823
QY 751 -----LVEATLEESACLSDLVHPE 769
Db 824 LLLQVHDELILEAPKEIERLCRLV-PE 850

RESULT 13
US-08-394-232A-32
; Sequence 32, Application US/08394232A
; Patent No. 6100078
; GENERAL INFORMATION:
; APPLICANT: RIGGS, MICHAEL G.
; APPLICANT: SIVARAM, MATHOOR
; APPLICANT: TUDOR, STARLA D.
; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS STEAROTHERMOPHILU
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gen-Probe Incorporated
; STREET: 9880 Campus Point Drive
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

```

; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA: US/08/394,232A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/307,410
; FILING DATE: 16-SEP-1994
; APPLICATION NUMBER: 08/222,612
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fisher, Carlos A
; REGISTRATION NUMBER: 36,510
; REFERENCE/DOCKET NUMBER: GP94003.CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-2807
; TELEFAX: 619-546-7929
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 876 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; US-08-394-232A-32
;
; Query Match 3.2%; Score 128.5; DB 2; Length 876;
; Best Local Similarity 21.0%; Pred. No. 0.0056;
; Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;
;
QY 99 HFLTDGIRAFARVSKPQFGSDDALISRRILQAACETLSNPRRRY-NEGLLDDE 157
DB 107 HYEADDIIGTMAAREE---GPAVKVIGSDRLTQ-----LASPQVTVEITKGGITDIE 158
QY 158 EATVITDV-----PMDKVPG-----ALCVLQEGGETEIVLRVG 190
DB 159 SYPTETVVEKYGLTPQIVDLKGLMGDKSDNIPEVPGIGEKTAVKLLKQGTVENVLASI 218
QY 191 EALLKERLPKSFQ--DVLVMAFAFDVSRDA--MALDPPDFITGYEFVEBALKLLQE 245
DB 219 DEIKGEKLENLRQYRDAL--LSQLAAICRDAPELTLD--DIVYKGEDEKVVAFQE 275
QY 246 EGASSLAPDLRAQIDE-----TLEETTPRVLELLGL---PLGDDYAARKRLNG- 290
DB 276 LGFQSFDRMAVQTDGEGKPLAGMDPAIADSVTDEMLADKAALVWEVVGDNVHAPIVGI 335
QY 291 -----LSGVRNILWSVGGGASALVGGLTREKFMNEAFLRWTAEE----- 330
DB 336 ALANERGRFLPETALADPKFLAW-----LGDETCKTMTFDSKAAVALKWKGI 385
QY 331 -----QVDLFVATPSNIPAES-----PEVVEVALALVAQAFIGK-----KPHLLQ 370
DB 386 ELRGVVFDLLLAAYLLDPAQAAGDVAAVAKHQYEAVRSDAEVYGGAKRTVPDEFTLAE 445
QY 371 DADKQFQQLQQAQVMAEIPAMLYDTRNWE-IDFGLERGLCALLIG-----KVDECRM 423
DB 446 HLAR-----KAAAIWALEBFLMDELRRNEQDRLLTELEQPLAGILANMEFTGVKVDTKKL 500
QY 424 -NLGLD-SEDSQYRNPAIVEFVLENSNRDNDLPLGLKLETLWLAGVVPFRDPTDKK 481
DB 501 EQMGAELETOLO-----AVERRIYELAGQEFNINSP---KQLGT-----VLPDKLQLPVLKK 549
QY 482 FKLGDYDDPMVLSYL-ERVEVVQGSPLAAATMARIAGAEHVKASAMQALQKVPFSRYTD 540
DB 550 TKYG-YSTADVLEKLAHPHEIVE--HILHYQLQKLGQSTYIE-GLLKVVHPVTGKVHTM 605
QY 541 RNSABPKVOQET--VFSVDVPGNVGRDGEPPGVFTAEAVRPS-----NFTNDYAI 590
;
; 606 FNQA-----LTQTGRLLSVPEPNLQNIPIRLEGRKIRQAFVPSEPDWLIFAADYSQIELRV 661
; 591 RAGVSESSVDETTVE-----MSVADMLKE---ASVKILAAAGVAIGLISLFSQ 634
; 662 LAHTAE---DDNLIEAPRRGLDIHTKTAMDIFHVSEEDVTANMRQAQAVNFGIVYGISD 718
; 635 KYFLKSSSSFOKDWVSSMESDVATIGSVRADDSEALPRMDARTAEINIVSKWKIKSLAF 694
; 719 -YGLAQNLNITRKEAAEFIERYPFASPPGVK-----QYMDNIVQE-AKOKGYVT 764
; 695 GPDHRIEMLPEVLDCRM-LKIWTDRAA--ETAQLGLVVDVTLTKLSVD-SVTYSADGTRA- 750
; 765 TLLHRRYLPDITTSRNFVRSFAETAMTPIQGSAAD-IKKAMIDLSVRLREERLQAR 823
; 751 -----LVEATLEESACLSDLVHPE 769
; 824 LLLQVHDELILEAPKEIERLCLRV-PE 850
;
; RESULT 14
; US-08-394-232A-34
; Sequence 34, Application US/08394232A
; Patent No. 6100078
; GENERAL INFORMATION:
; APPLICANT: RIGGS, MICHAEL G.
; APPLICANT: SIVARAM, MATHOOR
; APPLICANT: TUDOR, STARLA D.
; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS STEAROTHERMOPHILU
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gen-Probe Incorporated
; STREET: 9880 Campus Point Drive
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/394,232A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/307,410
; FILING DATE: 16-SEP-1994
; APPLICATION NUMBER: 08/222,612
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fisher, Carlos A
; REGISTRATION NUMBER: 36,510
; REFERENCE/DOCKET NUMBER: GP94003.CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-2807
; TELEFAX: 619-546-7929
; TELEX:
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 876 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; US-08-394-232A-34
;
; Query Match 3.2%; Score 128.5; DB 2; Length 876;
; Best Local Similarity 21.0%; Pred. No. 0.0056;
```


| | Matches | 170; | Conservative | 135; | Mismatches | 302; | Indels | 201; | Gaps | 431; |
|----|---------|---|--------------|------|------------|------|--------|------|------|------|
| Qy | 99 | HFUTDGIRRAFEARVSKPPQFGSDDALISRRQILOACETLSNPRSRREY-NEGLLDDE | 157 | | | | | | | |
| Db | 107 | HYEADDIIIGTMAARAERE---GFAVKVISGDRDLTQ---LASPQVTEITTKGITDIE | 158 | | | | | | | |
| Qy | 158 | EATVITDV-----PWKVPQ-----ALCVLOSGGETEIVLVRG | 190 | | | | | | | |
| Db | 159 | SYTPTVEKYGTTPQIVDLKGLMGDKSNIPGVPGIGKTAVKLLKGQGTVENVLASI | 218 | | | | | | | |
| Qy | 191 | EALLKERLPKSFQK--DVLVLMAFAFDVSRDA---MALDPDPDITGYEFVEEAKLLQE | 245 | | | | | | | |
| Db | 219 | DEIKGEKLENLRQYRDIAL-LSQLAAICRDAPVELTLD--DIVYKGEDREKVVALFQE | 275 | | | | | | | |
| Qy | 246 | EGASSIAPDLRAQIDE-----TLBEITPRYVLELLGL---PLGDDYAARKLNG- | 290 | | | | | | | |
| Db | 276 | LGFQSFLDKMAVQTDEGEKPLAGMDFAIADSVTDEMLADKAALVVEVVGVDNYHAPIVGI | 335 | | | | | | | |
| Qy | 291 | -----LSGVNRNLTWSVGGGASALVGLTREKPMNEAFLRMTAAE---330 | | | | | | | | |
| Db | 336 | ALANERGRFFLRPETALADPKFLAW-----LGDETKYKTMFDSKRAAVALKWKGI | 385 | | | | | | | |
| Qy | 331 | -----QVDLFVATPSNIPAES-----FEYEVVALALVAQAFICK-----KPHLLQ | 370 | | | | | | | |
| Db | 386 | ELRGVWFDLLLAAYLLDPAQAAGDVAAVAAMQHYEAVRSDEAVYKGAKRTVPDEPTLAE | 445 | | | | | | | |
| Qy | 371 | DADQKFOOLOQAQKVMAMEIPAMLYDITANNWE-IDFGLERGLCALLIG-----KVDSERM | 423 | | | | | | | |
| Db | 446 | HLAR-----KAAAIWALEEPIMDELNRNEODRLLEQPLAGILANMEFTGVKVDTKRL | 500 | | | | | | | |
| Qy | 424 | -WLGLD-SEDSQYRNPAIVFVLENSNRDNDLDLPLGLKLIETWLAGVFPFRDTRDKKX | 481 | | | | | | | |
| Db | 501 | EQMGAELTEQLQ---AVERRIYELAQEFNINSP---KQLGT---VLFDKQLQVLVKX | 549 | | | | | | | |
| Qy | 482 | FKLGDYDDPMWLSYL-ERVEVVGQSPAAAAWTMARIGAEHVKASAMQALQKVPSPRYTD | 540 | | | | | | | |
| Db | 550 | TKTG-YSTSADVLEKLAPHHEIVE--HILHYRQLGKLQSTYIE-GLLKVHPVPTGKVHTM | 605 | | | | | | | |
| Qy | 541 | RNSAEPKQVQET--VFSVDVPGNVNVRDGPVFIAEAVRSE-----NEETNDYAI | 590 | | | | | | | |
| Db | 606 | FNQA---LTQTRLSSVEPNLQNIPIRLBGRKIQRAFVSEPDWLIIPADYISQIEIURV | 661 | | | | | | | |
| Qy | 591 | RAGVSSESVDETVE-----MSVADMLKE---ASVKILAAAGVAIGLISLFSQ | 634 | | | | | | | |
| Db | 662 | LAHTAE---DONLIEAFRRGLDIHTKTAMDIFHVSEEDVTANMRROAKAVNFGIVYGISD | 718 | | | | | | | |
| Qy | 635 | KYFLKSSSSFORQDMVSSMESDVATIGSVRADDSREALPRMDARTAEINIVSKWQIKSLAF | 694 | | | | | | | |
| Db | 719 | -YGLAQNLTNTRKEAAEFIERYPASFPGVK-----QYMDNIVQOE-AKQKGYVT | 764 | | | | | | | |
| Qy | 695 | GPDHRIMBLEPVLDRGM-LKIWTDRAA-ETAAQLGLVYDYTLTKLSVD-SVTVSADGTRA- | 750 | | | | | | | |
| Db | 765 | TLLHRRYLPDITSRNFNVRSFAERTAMNTPIQSSAAD-IKKAMIDLSVTLRERLQAR | 823 | | | | | | | |
| Qy | 751 | -----LVEATPLESACLSDLVHPE | 769 | | | | | | | |
| Db | 824 | LLLOVHDELILEAPKEIERLCRIV-PE | 850 | | | | | | | |

```

RESULT 15
PCT-US95-04080-20
; Sequence 20, Application PC/TUS9504080
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS
; TITLE OF INVENTION: STEAROTHERMOPHILUS
; NUMBER OF SEQUENCES: 34
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EFO)
; CURRENT APPLICATION DATA:

```

Search completed: December 4, 2005, 12:29:59

Job time : 51 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 12:39:31 : Search time 233 Seconds

(without alignments)
2425.443 Million cell updates/sec

Title: US-10-600-070B-2

Perfect score: 4063

Sequence: 1 MEALSHVGICLSFQLCRUP.....YEVFWSKSGWKITEGSLAS 801

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--------------|
| 1 | 4063 | 100.0 | 801 | 2 | Q7XAS0 ARATH |
| 2 | 4052 | 99.7 | 801 | 2 | Q7XAR9 ARATH |
| 3 | 4051 | 99.7 | 801 | 2 | Q9FTG9 ARATH |
| 4 | 1775.5 | 43.7 | 760 | 2 | Q6Z729 ORYSA |
| 5 | 1775.5 | 43.7 | 760 | 2 | Q7PC78 ORYSA |
| 6 | 1646 | 40.5 | 324 | 2 | Q7XAS1 ARATH |
| 7 | 497 | 12.2 | 798 | 2 | Q8YTL0 ANASP |
| 8 | 448.5 | 11.0 | 626 | 2 | Q7NFP3 GLOVI |
| 9 | 442 | 10.9 | 631 | 2 | Q93AK0 SYNPF |
| 10 | 442 | 10.9 | 631 | 2 | Q5N029 SYNPF |
| 11 | 431 | 10.6 | 819 | 2 | Q8VY16 ARATH |
| 12 | 429 | 10.6 | 714 | 2 | Q55559 SYNPF |
| 13 | 425 | 10.5 | 841 | 2 | Q7L0L2 ARATH |
| 14 | 423 | 10.4 | 673 | 2 | Q8DKU7 SYNEL |
| 15 | 344.5 | 8.5 | 702 | 2 | Q7V0H1 PROMP |
| 16 | 340 | 8.4 | 653 | 2 | Q7U5S7 SYNPF |
| 17 | 318 | 7.8 | 685 | 2 | Q7VNU3 PROMA |
| 18 | 317 | 7.8 | 661 | 2 | Q7V8J7 PROMM |
| 19 | 155.5 | 3.8 | 5362 | 2 | Q93155 BACSU |
| 20 | 152.5 | 3.8 | 1263 | 2 | Q5KIT6 CRYNE |
| 21 | 148.5 | 3.7 | 5363 | 2 | Q7QJ29 BACAM |
| 22 | 148 | 3.6 | 5363 | 2 | Q6VK40 BACSU |
| 23 | 145 | 3.6 | 2297 | 2 | Q8YL08 ANASP |
| 24 | 143 | 3.5 | 2089 | 2 | Q39478 GSTRA |
| 25 | 141.5 | 3.5 | 1329 | 2 | Q4IA74 GIBZE |
| 26 | 141 | 3.5 | 2076 | 2 | Q6LZY0 METMP |
| 27 | 140.5 | 3.5 | 1835 | 2 | Q6BMT4 DEBHA |
| 28 | 139 | 3.4 | 1027 | 2 | Q7RYV6 NEUCR |
| 29 | 138 | 3.4 | 1274 | 2 | Q5STK0 CRYNE |
| 30 | 136.5 | 3.4 | 1954 | 2 | Q7MIT8 VIBVU |
| 31 | 134 | 3.3 | 678 | 2 | Q8PHZ9 XANAC |
| 32 | 134 | 3.3 | 1238 | 2 | Q9LEP9 BRANA |
| 33 | 133 | 3.3 | 1241 | 2 | Q9UKK5 HUMAN |
| 34 | 132.5 | 3.3 | 1011 | 2 | Q93GJ3 GRICK |
| 35 | 132.5 | 3.3 | 1951 | 2 | Q8DB34 VIBVU |
| 36 | 132.5 | 3.3 | 2111 | 1 | MCAS MYCBO |
| 37 | 132 | 3.2 | 926 | 2 | Q4JUH6 CORJK |
| 38 | 132 | 3.2 | 3213 | 2 | Q4HMX8 GIBZE |
| 39 | 131.5 | 3.2 | 1194 | 1 | DFOL VZVD |
| 40 | 131.5 | 3.2 | 1194 | 2 | Q6QCM7 HHV3 |
| 41 | 131.5 | 3.2 | 2111 | 2 | P96291 MYCTU |
| 42 | 131 | 3.2 | 852 | 2 | Q6BSS1 DEBHA |
| 43 | 130.5 | 3.2 | 876 | 1 | DPO1 BACST |
| 44 | 130.5 | 3.2 | 1009 | 2 | Q93G19 RICCN |
| 45 | 130.5 | 3.2 | 1009 | 2 | Q93GJ2 RICCN |

RESULT 1

Q7XAS0 ARATH

ID Q7XAS0 ARATH PRELIMINARY; PRT; 801 AA.

AC Q7XAS0;

DT 01-OCT-2003 (TRENBLrel. 25, Created)

DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)

DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)

DE Division protein.

GN Name=ARC6;

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

OX NCBI_TaxID=3702;

RN [1]

RP NUCLEOTIDE SEQUENCE

RA MEDLINE=22779057; Pubmed=12897262; DOI=10.1105/tpc.013292;

RA Vitha S., Froehlich J.E., Koksharova O., Pyke K.A., Van Erp H.,

RT OsterYoung K.W.;

RT "ARC6 is a J-domain plastid division protein and an evolutionary

RT descendant of the cyanobacterial cell division protein Ftn2.";

RL Plant Cell 15:1918-1933(2003).

DR EMBL; AY221468; AAQ18645.1; -; Genomic DNA.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0031072; F:heat shock protein binding; IEA.

DR GO; GO:0051082; F:unfolded protein binding; IEA.

DR GO; GO:0006457; P:protein folding; IEA.

DR InterPro; IPR001623; DnaJ N.

SQ SEQUENCE 801 AA; 88306 MW; D09E27253E46FCC6 CRC64;

Query Match 100.0%; Score 4063; DB 2; Length 801;

Best Local Similarity 100.0%; Pred. No. 4.7e-250;

Matches 801; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEALSHVGICLSFQLCRLLPATTKLRSNHTSTTICSAKWADRLLDNFNFTSSSSSS 60

Db 1 MEALSHVGICLSFQLCRLLPATTKLRSNHTSTTICSAKWADRLLDNFNFTSSSSSS 60

Qy 61 FATATTATLVSPPSDRPERHVPIPIDFVVLGAOTHFDTGIRRAFEARVSKPQFG 120

Db 61 FATATTATLVSPPSDRPERHVPIPIDFVVLGAOTHFDTGIRRAFEARVSKPQFG 120

Qy 121 FSDDALISRRQILQAACETLSNPRSRREYNEGLDDEATVITDVPMDKVPALCVLQEG 180

Db 121 FSDDALISRRQILQAACETLSNPRSRREYNEGLDDEATVITDVPMDKVPALCVLQEG 180

Qy 181 GETEIVLRVGEALLKERLPKSFQDQVVVLVMAFLDVSRDAMALDPPDFTGYEFVEEAL 240

Db 181 GETEIVLRVGEALLKERLPKSFQDQVVVLVMAFLDVSRDAMALDPPDFTGYEFVEEAL 240

Qy 241 KLLQEGAGSLAPDLRAQIDETLEEITPRVYVLELLGLPLGDDYAAKRLNGLSGVRNIIWS 300

Db 241 KLLQEGAGSLAPDLRAQIDETLEEITPRVYVLELLGLPLGDDYAAKRLNGLSGVRNIIWS 300

```
QY 301 VGGGASALVGGITREKFNEAFLRMTAAEOVDL FVATPSNIPAESFEYVEVALVAQA 360
Db 301 VGGGASALVGGITREKFNEAFLRMTAAEQVDL FVATPSNIPAESFEYVEVALVAQA 360
QY 361 FIGKPHLLQDADKQFOQLQQAQKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
Db 361 FIGKPHLLQDADKQFOQLQQAQKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
QY 421 CRMWLGDSSEDSQYRNPAIVEFVLENSNRDNDLPLGCKLLETWLAGVVPFRDRTDK 480
Db 421 CRMWLGDSSEDSQYRNPAIVEFVLENSNRDNDLPLGCKLLETWLAGVVPFRDRTDK 480
QY 481 KFKLGDDYDDPMVLSYLERVEVVGSGPLAAATMARI GAHVKASAMQALQKVPFRYTD 540
Db 481 KFKLGDDYDDPMVLSYLERVEVVGSGPLAAATMARI GAHVKASAMQALQKVPFRYTD 540
QY 541 RNSAEPKDVQETVFSVDPVGNVGRDGEPCGVFAEAVRPSNPENFTNDYAIRAGVSSSVD 600
Db 541 RNSAEPKDVQETVFSVDPVGNVGRDGEPCGVFAEAVRPSNPENFTNDYAIRAGVSSSVD 600
QY 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFSQKYLKSSSSFORQDMVSSMESDVATI 660
Db 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFSQKYLKSSSSFORQDMVSSMESDVATI 660
QY 661 GSVRADDSSEALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAA 720
Db 661 GSVRADDSSEALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAA 720
QY 721 ETAQLGLVVDYTLILKLSVDSVTSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
Db 721 ETAQLGLVVDYTLILKLSVDSVTSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
QY 781 RYEVFWSKSGWKITEGSLVAS 801
Db 781 RYEVFWSKSGWKITEGSLVAS 801

RESULT 2
Q7XAR9 ARATH
ID Q7XAR9 ARATH PRELIMINARY; PRT; 801 AA.
AC Q7XAR9
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Division protein.
GN Name=ARC6;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22779057; PubMed=12897262; DOI=10.1105/tpc.013292;
RA Vitha S., Froehlich J.E., Koksharova O., Pyke K.A., Van Erp H.,
RA Osteryoung K.W.;
RT "ARC6 is a J-domain plastid division protein and an evolutionary
RT descendant of the cyanobacterial cell division protein Ftn2.";
RL Plant Cell 15:1918-1933(2003).
DR EMBL; AY221469; AAQ18646.1; -; mRNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0031072; F:heat shock protein binding; IEA.
DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR GO; GO:0006457; P:protein folding; IEA.
DR InterPro; IPR01623; DnaJ_N.
SQ SEQUENCE 801 AA; 88248 MW; 7E21B3FD4BE4B61 CRC64;

Query Match 99.7%; Score 4052; DB 2; Length 801;
Best Local Similarity 99.8%; Pred. No. 2 4e-249;
Matches 799; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MEALSHVGIGLSPFQLCRLPPATTKLRRSHNTSTTICSAKWADRLLSDFNFTSDSSSS 60
Db 1 MEALSHVGIGLSPFQLCRLPPATTKLRRSHNTSTTICSAKWADRLLSDFNFTSDSSSS 60
QY 61 FATATTATLVSPPSIDRPERHVPIPIDFYVQLGAQTHFLTDGIRRAFEARVSKPPQFG 120
Db 61 FATATTATLVSPPSIDRPERHVPIPIDFYVQLGAQTHFLTDGIRRAFEARVSKPPQFG 120
QY 121 FSDDALISRRQIILOAACETLSNPRSRREYNELGLDDEEATVITDVPMDKVPKALCVLQEG 180
Db 121 FSDDALISRRQIILOAACETLSNPRSRREYNELGLDDEEATVITDVPMDKVPKALCVLQEG 180
QY 181 GETEIVLRVGEALLKERLPKSFQDQVVLVMAALFLDVSRDAMALDPPDFTTGYEFVEBAL 240
Db 181 GETEIVLRVGEALLKERLPKSFQDQVVLVMAALFLDVSRDAMALDPPDFTTGYEFVEBAL 240
QY 241 KLAQEGASSLAPDLRAQIDETLEEITPRVYVLELLGLPLGDDYAAKRLNGLSGVRNLIWS 300
Db 241 KLAQEGASSLAPDLRAQIDETLEEITPRVYVLELLGLPLGDDYAAKRLNGLSGVRNLIWS 300
QY 301 VGGGASALVGGITREKFNEAFLRMTAAEQVDL FVATPSNIPAESFEYVEVALVAQA 360
Db 301 VGGGASALVGGITREKFNEAFLRMTAAEQVDL FVATPSNIPAESFEYVEVALVAQA 360
QY 361 FIGKPHLLQDADKQFOQLQQAQKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
Db 361 FIGKPHLLQDADKQFOQLQQAQKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420
QY 421 CRMWLGDSSEDSQYRNPAIVEFVLENSNRDNDLPLGCKLLETWLAGVVPFRDRTDK 480
Db 421 CRMWLGDSSEDSQYRNPAIVEFVLENSNRDNDLPLGCKLLETWLAGVVPFRDRTDK 480
QY 481 KFKLGDDYDDPMVLSYLERVEVVGSGPLAAATMARI GAHVKASAMQALQKVPFRYTD 540
Db 481 KFKLGDDYDDPMVLSYLERVEVVGSGPLAAATMARI GAHVKASAMQALQKVPFRYTD 540
QY 541 RNSAEPKDVQETVFSVDPVGNVGRDGEPCGVFAEAVRPSNPENFTNDYAIRAGVSSSVD 600
Db 541 RNSAEPKDVQETVFSVDPVGNVGRDGEPCGVFAEAVRPSNPENFTNDYAIRAGVSSSVD 600
QY 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFSQKYLKSSSSFORQDMVSSMESDVATI 660
Db 601 ETTVMSVADMLKEASVKILAAAGVAIGLISLFSQKYLKSSSSFORQDMVSSMESDVATI 660
QY 661 GSVRADDSSEALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAA 720
Db 661 GSVRADDSSEALPRMDARTAEINIVSKWKIKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAA 720
QY 721 ETAQLGLVVDYTLILKLSVDSVTSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
Db 721 ETAQLGLVVDYTLILKLSVDSVTSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
QY 781 RYEVFWSKSGWKITEGSLVAS 801
Db 781 RYEVFWSKSGWKITEGSLVAS 801

RESULT 3
Q9FIG9 ARATH
ID Q9FIG9 ARATH PRELIMINARY; PRT; 801 AA.
AC Q9FIG9
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Arabidopsis thaliana genomic DNA, chromosome 5, pl clone: MDH9
DE (Hypothetical protein At5g42480).
GN Name=At5g42480;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99156233; PubMed=10048488;
```

| | |
|--|--|
| RA | Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N., |
| RA | Tabata S.; |
| RT | "Structural analysis of Arabidopsis thaliana chromosome 5. VIII. |
| RT | Sequence features of the regions of 1,081,958 bp covered by seventeen |
| RT | physically assigned P1 and TAC clones."; |
| RL | DNA Res. 5:379-391(1998). |
| RN | [2] |
| RP | NUCLEOTIDE SEQUENCE. |
| RA | Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M., |
| RA | Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L., |
| RA | Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L., |
| RA | Carinci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., |
| RA | Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J., |
| RA | Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J., |
| RA | Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K., |
| RA | Davis R.W., Ecker J.R., Theologis A.; |
| RL | Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases. |
| RN | [3] |
| RP | NUCLEOTIDE SEQUENCE. |
| RA | Yamada K., Chan M.M., Chang C.H., Dale J.M., Heuan V.W., Lee J.M., |
| RA | Quach H.L., Tang C., Toriumi M., Wallender E.K., Wong C., Wu H.C., |
| RA | Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y., |
| RA | Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M., |
| RA | Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., |
| RA | Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R., |
| RA | Theologis A.; |
| RL | Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases. |
| RL | EMBL; AB016888; BAB10489.1; -; Genomic_DNA. |
| DR | EMBL; AY091075; AAM13895.1; -; mRNA. |
| DR | EMBL; AY150490; AAN12907.1; -; mRNA. |
| DR | GO: GO:0016021; C: integral to membrane; IEA. |
| DR | GO: GO:0031072; F: heat shock protein binding; IEA. |
| DR | GO: GO:0051082; F: unfolded protein binding; IEA. |
| DR | GO: GO:0006457; P: protein folding; IEA. |
| DR | InterPro; IPR001623; DnaJ_N. |
| KW | Hypothetical protein. |
| SQ | SEQUENCE 801 AA; 88260 MW; 608E776FBA73FECF CRC64; |
| Query Match 99.7%; Score 4051; DB 2; Length 801; | |
| Best Local Similarity 99.8%; Pred. No. 2.7e-249; | |
| Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0; | |
| QY | 1 MEALSHVIGIGLSPQLCRLPATTKLRRSHNTSTTICSASKWADRLLSDFNFTSDSSSS 60 |
| DB | 1 MEALSHVIGIGLSPQLCRLPATTKLRRSHNTSTTICSASKWADRLLSDFNFTSDSSSS 60 |
| QY | 61 FATATTATTVLSLPPSIDRPERHVPIDIFQVILGAQTHFLTDGIRRAFEARVSKPPQFG 120 |
| DB | 61 FATATTATTVLSLPPSIDRPERHVPIDIFQVILGAQTHFLTDGIRRAFEARVSKPPQFG 120 |
| QY | 121 FSDDALISRQILQAACETLSNPRSREYNEGLLDDEATVITDVPMDKVPALCVLQEG 180 |
| DB | 121 FSDDALISRQILQAACETLSNPRSREYNEGLLDDEATVITDVPMDKVPALCVLQEG 180 |
| QY | 181 GETEIVLRVEALLKERLPKSFQDVVLVWALAFLDVSRDAMALDPPDFTGYEFVEAL 240 |
| DB | 181 GETEIVLRVEALLKERLPKSFQDVVLVWALAFLDVSRDAMALDPPDFTGYEFVEAL 240 |
| QY | 241 KLQEEGASSLPDLRAQIDETLEEITPRYVLELLGLPLGDDYAAKRLNGLSGVRNIIWS 300 |
| DB | 241 KLQEEGASSLPDLRAQIDETLEEITPRYVLELLGLPLGDDYAAKRLNGLSGVRNIIWS 300 |
| QY | 301 VGGGASALVGGLTREKFMNEAFLRMTAAEOVDLFVATPSNIPAESFEVVEVALVAQA 360 |
| DB | 301 VGGGASALVGGLTREKFMNEAFLRMTAAEOVDLFVATPSNIPAESFEVVEVALVAQA 360 |
| QY | 361 FIGKPKHLLQADKQFOOLQAQKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420 |
| DB | 361 FIGKPKHLLQADKQFOOLQAQKVMAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVD 420 |
| QY | 421 CRMWGLGDSQVRNPDAI VEFVLNENRDNDDLPGLCKLLETWLAVGVPPFRDFTKOK 480 |
| DB | 421 CRMWGLGDSQVRNPDAI VEFVLNENRDNDDLPGLCKLLETWLAVGVPPFRDFTKOK 480 |

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]

NUCLEOTIDE SEQUENCE.

RX MEDLINE=22779057; PubMed=12897262; DOI=10.1105/tpc.013292;
 RA Vitha S., Froehlich J.E., Koksharova O., Pyke K.A., Van Erp H.,
 RA Osteryoung K.W.;
 RA "ARC6 is a J-domain plastid division protein and an evolutionary
 RT descendant of the cyanobacterial cell division protein Ftn2.";
 RL Plant Cell 15:1918-1933(2003).

DR EMBL; AY221467; AAQ18644.1; -; Genomic DNA.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0031072; F:heat shock protein binding; IEA.

DR GO; GO:0051082; F:unfolded protein binding; IEA.

DR GO; GO:0006457; P:protein folding; IEA.

DR InterPro; IPR001623; DnaJ_N.

SQ SEQUENCE 324 AA; 35468 MW; 0CC5C0CC469A6325 CRC64;

Query Match 40.5%; Score 1646; DB 2; Length 324;
 Best Local Similarity 100.0%; Pred. No. 1.4e-96;
 Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEALSHVIGLSPPQLCLRPATTKLRSHNTSTTICSAKWADRLLSDFNTSDSSSS 60
 DB 1 MEALSHVIGLSPPQLCLRPATTKLRSHNTSTTICSAKWADRLLSDFNTSDSSSS 60
 QY 61 FATATTATLVSPPSIDRPERHVPIDFQVLGAQTHFTDGRRAFEARVSKPPQFG 120
 DB 61 FATATTATLVSPPSIDRPERHVPIDFQVLGAQTHFTDGRRAFEARVSKPPQFG 120
 QY 121 FSDALISRQILQAACETLSNRSREYNEGLDDEATVITDVPWPKVPGALCVLQEG 180
 DB 121 FSDALISRQILQAACETLSNRSREYNEGLDDEATVITDVPWPKVPGALCVLQEG 180
 QY 181 GETEIVLRGEALLKERLPKSFQDVVLVMALELDVSRDAMALDPDFITGYEFVEAL 240
 DB 181 GETEIVLRGEALLKERLPKSFQDVVLVMALELDVSRDAMALDPDFITGYEFVEAL 240
 QY 241 KLLQEGGASSLPDLRAQIDETLEETPRVYVLELLGLPLGDDYAAKRLNGLSGVNLIWS 300
 DB 241 KLLQEGGASSLPDLRAQIDETLEETPRVYVLELLGLPLGDDYAAKRLNGLSGVNLIWS 300
 QY 301 VGGGASALVGGITREKFMNEAFL 324
 DB 301 VGGGASALVGGITREKFMNEAFL 324

RESULT 7

Q8YTL0 ANASP PRELIMINARY; PRT; 798 AA.
 AC Q8YTL0;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE All2707 protein.
 GN OrderedLocusNames=all2707;
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]

NUCLEOTIDE SEQUENCE.

RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RA "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium Anabaena sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).
 DR EMBL; BA000019; BAB74406.1; -; Genomic_DNA.
 DR FIR; AD2144; AD2144.

DR HSSP; P08622; 1B0Z.
 DR GO; GO:0031072; F:heat shock protein binding; IEA.
 DR GO; GO:0051082; F:unfolded protein binding; IEA.
 DR GO; GO:0006457; P:protein folding; IEA.
 DR InterPro; IPR001623; DnaJ_N.

DR Pfam; PF00226; DnaJ_1.

DR SMART; SM00271; DnaJ_1.

DR PROSITE; PS50076; DnaJ_2; 1.

DR Chapterone; Complete proteome.

SQ SEQUENCE 798 AA; 90124 MW; 3C57B66C221722AA CRC64;

Query Match 12.2%; Score 497; DB 2; Length 798;

Best Local Similarity 24.6%; Pred. No. 9.6e-23;

Matches 207; Conservative 139; Mismatches 301; Indels 196; Gaps 33;

QY 84 VPIPIDFYQVLGAQTHFTDGRRAFEARVSKPPQFGFSDALISRRQILQAACETLSNP 143

DB 11 VRIPLDYRILGLPLAASDEQLRQAYSDRIVQLPRREYSQAALSRKQLIEEAYVVLSDP 70

QY 144 RSRREY-----NEGLLDDEEATVITDVPWPKVPGALCV 176

DB 71 KERSYDOLYLALAHAYDPDAAATTKVAENRGDSNNGHFDVQSLSI--EVSSEBELIGALLI 128

QY 177 LOEGGETEIVLRVGEALL-----KEELPKSF-----KQDVLMALAFIDVSR 219

DB 129 LOELGEYELVLKLGRLNYLGNQNGTASTRNGNRTPEEFLDSSERPDIILTVALASLELGR 188

QY 220 D-----AMALDPPDFITGYEFVEEALKLLQEGGASSLPDLRAQIDETLEETPR 269

DB 189 EQWQGHVENAALSLE-----TGQE-----VLPSEG---IFPSVQAEIQADLYKLRPY 233

QY 270 YVLELLGLPLGDDYAAKRLNGLSGVNLIWSVGG--GGASALVGGITRE---KPMNEAFLR 325

DB 234 RILELLALP--QEKTIERHQGLDLQLSILDDRGIGDTGTNDQSLNIDFLRFIQQLRHH 291

QY 326 MTAEOVDLFAVTSNIPAESFEVVEVALALVAQAFIGKKPHLLQADKQFOOLOAKVM 385

DB 292 LTVAEQHKLFDG--ESKRPS--AVATYLVAYASTARTGTQRPALIRHAKQILMLRSKQ-- 347

QY 386 AMEIPAMLYDRNNWEIDFGLERGLCALLIGKVDRCMWLGDSDSQVNRNPAIVEFVLE 445

DB 348 -----DVHLEQSLCALLLQTEATEATRVLELSQE---YEALALI-----R 393

QY 446 NSNRDNDLFLGCKLLETLAGVVFPRFDTKKKFKLGDYDDPMVLSYLERV----- 500

DB 384 EKSQDSPDLLPGLCLYAEQWLQNEVFPHFRLDSRQASLKDYFANQVQVQAYLEALPDAE 443

QY 501 -----EVVQGSPLAAATMAR--IGAHE--VKASAMQALQKVFPS---RYTD 540

DB 444 TTNEWAVINRQSFQSGNYSGGTTPVAKRPVGKARRPGEASTRPVQRPQSHPEVNRQPH 503

QY 541 RNSAEPKDVQETVFSVDVPGNNVGRDGE-----PGVFIAEAVRPESENFETND 587

DB 504 QNRTDPELPETSNHRRPESNFTTARENISTTDAYTNYPEIPVERASRP----- 555

QY 588 YAIRAGVS---ESSVDETTV-----EMSVADMKEASVKILAA 622

DB 556 --VQPGVSGYQTOSTPPRQTPKRRKKPQAVNRGHSIHQQQPSPTLGRKTRLLWIVL 613

QY 623 GVAIGLISLF---SQKY-FLKS-----SSSFQKMDVMSSMESDVATIGSVRADDSEALPR 673

DB 614 G-SLGGILLFWLIVSTTFGLMKNVFFPAPSLQGEQLSIQISQPPLEIPKNAQIQSPEVS 672

QY 674 MDARTAEINYSKWQKIKSLAFGPDRRIEMLEVLVDGRMLKIWTDRAAETAQGLVDYDYL 733

DB 673 LITEARTKIENWLATKASALGAEHKIESLEIITGSALSQWRLIALQADKNRREYS- 731

QY 734 LKLSVDSVTVS--ADGTRALVEATEESACLSDLVH--PENNATDVRTTTRVEVFWMSKG 790

DB 732 HSKVDSISKSDIDPNRASVGTATVRE---LTQFYENGQKGKSDER-LRVYELIRQDDI 787

QY 791 WKI 793

[illegible]

```
Query Match      10.9%; Score 442; DB 2; Length 631;
Best Local Similarity 24.5%; Pred. No. 2.1e-19;
Matches 189; Conservative 118; Mismatches 262; Indels 204; Gaps 30;

QY 86 IPIDFYVLGAQTHFLTDGIRRAFEARVSKPPQFGSDALISRRQILQAACETLSNPRS 145
DB 3 IPLDYRILCVGVQASADKLAESYDRNLNQSFSHEFSELQAARRQLLEAAIAELSDPEQ 62
QY 146 RREYNE-----GLDDDEAVITVDPWDKVPKALCVLOEGGETEIVLRVGEALL----- 194
DB 63 RDRYDRFFQGLEAIEPSLELED--WQRI--GALLILLELGEYDRVVSQALBELLPDYDAS 119
QY 195 KERLPKSFQDQVILVMALAFLDVSRDAMALDPPDFTGYEFVEEALKLQEGE----- 248
DB 120 AEVRDQFARGDIALAIALSQSLGRCRQ-----OGLYEQAQHFGRS 162
QY 249 -SSLA-----PDLRAQIDETLEETTPRVYLELGLPLGDDYAAKRLNGLSGVRNII--W 299
DB 163 QSALADHQRPELSRTLHQEGQQLRPYRILERLAQPLTAD--SDRQQGLLLQLQMLDDRQ 220
QY 300 SVGGGGASALVGLTRKFKMNEAFRL-----MTAAEQVDLF---VATPSNIPAESFEVYE 351
DB 221 GIEGPGDGG--SGITLDNFI--MFLQIIRGVLTIAEQQLLFESEARRPS--PAASF---F 271
QY 352 VALALVAQAFIGKPKPHLLQADKQFOOQQAQVMAEIPAMLYDTRNNWEIDFGLERGLC 411
DB 272 ACYTLIARGFCDHQPISLIRASLLHELKS-----RMDVHIEQAIA 312
QY 412 ALLIGKVDCEKMWGLDSED-----SOVRNPAIVEFVLENSNRDNDLPLGKLETLWLA 467
DB 313 SLLGQPEEAELI--VQSQDEETLSQIRALAQGEAL-----IVGLCRFTETWLA 360
QY 468 GVVFPRFRDTRDKKFKLGDYDDPWLVSILYERVVVQSGPLAAATWARIAGAEHVKASAM 527
DB 361 TKVFPDRDLKERTAPLQPYFDDPDVQTYLDAIVEL-----VAPP 432
QY 528 QALQKVPFSRYTRDNRSAEPKDVQSVFSDP-----VGNVGRDGEQGVFIAEAVRPS 580
DB 397 -----PSDLMTPLPVEPLVRSLLAKELPTPATPG-----VAPP 432
QY 581 ENFETNDYAIRAGVSSVDTEVNSVADMLKESVKILAAQVAIGLISLFSQKYLKS 640
DB 433 PRRRDRSRPARTAKRLPLPWIGLV-----VVVLGGTGV-----WAWRS 475
QY 641 SSSFORKDMVSSMESDVATIGSVRADDSALP-----RMDARTAEINVSQWOKIKS 691
DB 476 RSN-----STPTPPPPVQVTLPEAVPAPSPAPVTVALDRAQAEIVLQNLWLAQA 524
QY 692 LAFGPDRHIMLPEVLDRMLKIWTDRAAETA--QLGLVVDYTLKLSDVSVTSADGTR 749
DB 525 AALGPQYDRDLATVLTGEVLQWQGFSSQAANTQLTSQPDH--KLTVDSVQLSDGDQR 581
QY 750 ALVEATLESACLSDLVHPENNAVDRT-----YTRTEVFWSKSGWKITEGSLV 799
DB 582 AVVQAKVDE-----VEQYTRGQDLLETRFDLGLVIRYQLVRENINWIKIASISLV 630

RESULT 10
QSN029_SYNP6 PRELIMINARY; PRT; 631 AA.
AC QSN029;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Cell division protein Ftn2 homolog.
GN OrderedLocustNames=sync2151 d;
OS Synchococcus sp. (strain PCC 6301) (Anacystis nidulans).
OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.
OX NCBI_TaxID=269084;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sugita M.;
RT "Complete genome structure of the unicellular cyanobacterium Anacystis
```

```
RT nidulans 6301 (Synchococcus sp. PCC6301)";
RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Participates actively in the response to hyperosmotic
CC and heat shock by preventing the aggregation of stress-denatured
CC proteins and by disaggregating proteins, also in an autonomous,
CC dnaK-independent fashion. Unfolded proteins bind initially to
CC dnaJ; upon interaction with the dnaJ-bound protein, dnaK
CC hydrolyzes its bound ATP, resulting in the formation of a stable
CC complex. GrpE releases ADP from dnaK; ATP binding to dnaK triggers
CC the release of the substrate protein, thus completing the reaction
CC cycle. Several rounds of ATP-dependent interactions between dnaJ,
CC dnaK and grpE are required for fully efficient folding. Also
CC involved, together with dnaK and grpE, in the DNA replication of
CC plasmids through activation of initiation proteins (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- DOMAIN: The J domain is necessary and sufficient to stimulate dnaK
CC ATPase activity. Zinc center 1 plays an important role in the
CC autonomous, dnaK-independent chaperone activity of dnaJ. Zinc
CC center 2 is essential for interaction with dnaK and for dnaJ
CC activity (By similarity).
DR EMBL: AP008231; BAD80341.1; -; Genomic DNA.
DR GO: GO:0031072; F:heat shock protein binding; IEA.
DR GO: GO:0051082; F:unfolded protein binding; IEA.
DR GO: GO:0051301; P:cell division; IEA.
DR GO: GO:0006457; P:protein folding; IEA.
DR InterPro: IPR001623; DnaJ_N.
DR Pfam: PF00226; DnaJ_1.
DR PROSITE: PS00076; DnaJ_2; 1.
DR Cell division; Chaperone; Complete proteome; DNA replication;
KW Heat shock.
SQ SEQUENCE 631 AA; 70694 MW; 6B46142E3B48DEE2 CRC64;

Query Match      10.9%; Score 442; DB 2; Length 631;
Best Local Similarity 24.5%; Pred. No. 2.1e-19;
Matches 189; Conservative 118; Mismatches 262; Indels 204; Gaps 30;

QY 86 IPIDFYVLGAQTHFLTDGIRRAFEARVSKPPQFGSDALISRRQILQAACETLSNPRS 145
DB 3 IPLDYRILCVGVQASADKLAESYDRNLNQSFSHEFSELQAARRQLLEAAIAELSDPEQ 62
QY 146 RREYNE-----GLDDDEAVITVDPWDKVPKALCVLOEGGETEIVLRVGEALL----- 194
DB 63 RDRYDRFFQGLEAIEPSLELED--WQRI--GALLILLELGEYDRVVSQALBELLPDYDAS 119
QY 195 KERLPKSFQDQVILVMALAFLDVSRDAMALDPPDFTGYEFVEEALKLQEGE----- 248
DB 120 AEVRDQFARGDIALAIALSQSLGRCRQ-----OGLYEQAQHFGRS 162
QY 249 -SSLA-----PDLRAQIDETLEETTPRVYLELGLPLGDDYAAKRLNGLSGVRNII--W 299
DB 163 QSALADHQRPELSRTLHQEGQQLRPYRILERLAQPLTAD--SDRQQGLLLQLQMLDDRQ 220
QY 300 SVGGGGASALVGLTRKFKMNEAFRL-----MTAAEQVDLF---VATPSNIPAESFEVYE 351
DB 221 GIEGPGDGG--SGITLDNFI--MFLQIIRGVLTIAEQQLLFESEARRPS--PAASF---F 271
QY 352 VALALVAQAFIGKPKPHLLQADKQFOOQQAQVMAEIPAMLYDTRNNWEIDFGLERGLC 411
DB 272 ACYTLIARGFCDHQPISLIRASLLHELKS-----RMDVHIEQAIA 312
QY 412 ALLIGKVDCEKMWGLDSED-----SOVRNPAIVEFVLENSNRDNDLPLGKLETLWLA 467
DB 313 SLLGQPEEAELI--VQSQDEETLSQIRALAQGEAL-----IVGLCRFTETWLA 360
QY 468 GVVFPRFRDTRDKKFKLGDYDDPWLVSILYERVVVQSGPLAAATWARIAGAEHVKASAM 527
DB 361 TKVFPDRDLKERTAPLQPYFDDPDVQTYLDAIVEL-----VAPP 432
QY 528 QALQKVPFSRYTRDNRSAEPKDVQSVFSDP-----VGNVGRDGEQGVFIAEAVRPS 580
DB 397 -----PSDLMTPLPVEPLVRSLLAKELPTPATPG-----VAPP 432
```

```
QY 581 ENFETNDYAIRAGVSESSVDETTVMNSVADMLKBAVSKILAAAGVAIGLISLFSQKYFLKS 640
Db 433 PRRRRDRSERPARTAKRLPLFWIGLV-----VVVLGGGTGV-----WAWRS 475
QY 641 SSSFORQKDWSSMESDVATIGSVRADDSALP-----RMDARTAEINIVSKWQKIIS 691
Db 476 RSN-----STPPPTPPPPVQTLPEAVPAPSPVTVVALDRAAQATVILQNLWLAAKA 524
QY 692 LAFGPDHRIEMLEPEVLDGRMLKIWTDRAAETA--QLGLVYDYTLKLSVDSVTSVSGTGR 749
Db 525 AALGPQYDRDLATVLTGEVLTQWGFSSQQAQNTQLTSQFDH---KLTVDVSQVLSGDQR 581
QY 750 ALVEATLESACLSLIVHPENNATVRT---YTTRYEVFWMSGKWKITGSQLV 799
Db 582 AVVQAKVDE---VEQVYRGDQLLETRDLGLVIRYQVLRENNINWIKIASISLV 630

RESULT 11
Q8VY16 ARATH PRELIMINARY; PRT; 819 AA.
AC Q8VY16;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein At3g19180.
GN Name:At3g19180;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.B., Toriumi M., Yu G., Bowser L.,
RA Carlinici P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY074283; AAL66980.1; -; mRNA.
KW Hypothetical protein.
SQ SEQUENCE 819 AA; 90750 MW; 6720083F4864657E CRC64;

Query Match 10.6%; Score 431; DB 2; Length 819;
Best Local Similarity 23.5%; Pred. No. 1.6e-18;
Matches 200; Conservative 116; Mismatches 311; Indels 224; Gaps 32;

QY 56 SSSSPATATTTATLVSLPPSIDRPRHVPPIPIIDFYQVLQAQTHFLTDGIRRAFEARVSK 115
Db 81 STTSLAATST-----TELPTCYQLIGVSEQAEDEVKSVINLKKT 124
QY 116 PPQGFSDDALISRRQILQAACETLSNPRSRREYNEGLDDDEATVITDVPMDKVPFALC 175
Db 125 DAEGYVTEAAAAQDQLMDVRDKL---LFSEYAGNLEKTIAPKSPLRIPAWLPGALC 181
QY 176 VLQEGGETEIVLRVBALLKERLPKSPKQDVVLVMAFLDVSRDAMALDPPDFTGYEF 235
Db 182 LLQEVQEKVLVDIGRAALRLNDSKPYIHDIFLSMALAECAIAKAEVFNKVS--QGPEA 239
QY 236 VEEALKLOEE-GASSLAPDLRAQIDETLEETIPRVVLELGLPLGDDVAAREKELNGLSV 294
Db 240 LARAQSLKSVTLGKLA--LTQIEESLEGAPPCTLDLGLPRTPEAERRRGAIAL 297
QY 295 RNILWSVGGGASALVGGITREK-----PMNEAFRLMTAAEQVDLF-----VATP 339
Db 298 RELLRQ-----GLSVEASCOIQDWPCFLSAISRLATEIVLDPDLDLAIKTR 346
QY 340 SNIPAESE-----VYEVALALVAQAFIKGPKHLLQDADKQFQLOQAQKVAWEIP 390
Db 347 NKKSLSESHNRVVIDPNCFFVLLGHIAVFGSG-----KQNETINKAKTICECLI 396
```

```
QY 391 AMLYDTRNNWEIDFGLERGLCALLIGKVDECRMWLGL-----DSDSQVRNPAIVEFVLEN 446
Db 397 A-----SEGVDLKFEAEFCFLKQGSAAELEKQLQLESNDSAVRNS-----ILGK 444
QY 447 SNRDDNDLPLCKLLETWLAGVVFPRFRDTK-----DKFKLGDYDDPMVL 494
Db 445 ESRSTS-----ATPSLEAWLMESVLANFPDPTRGCSPLANFFRAEKKYPENKMGSPSIM 499
QY 495 SY-----LERVEVQGSPLAAAAATMARICAEHVKASAMQALQKVPFSPRYTDRNSAEPKD 548
Db 500 NHKTNQRPLSTTQFVNS-----QHL-----YTAVEQLITFD 531
QY 549 VQTFVSVDPVGNVGRDGPVFIABAVRPSNFENFETNDYAIRAGVSESSV-DETTVMS 607
Db 532 LQSPVVSAK---NNDE-----TSASMPFSVLKRN-----LGVHKNKIWDEW---LS 571
QY 608 VADMLKEASVKILAAAGVAIGLISLFSQKY----- 636
Db 572 QSSLIGRVSVVAL-----LGCTVFFSLKLSGIRSLQSMPIVSARPHSESDSFLWKTE 626
QY 637 ---FLKSSSSFORQKDWSS-----MESDVATIGSVRADDSALP 673
Db 627 SGNFRKULSVNRNGIVGNIKVLIDMLKMGCEHPDLYLKSSQGSATLSHSASELHKR 686
QY 674 -MDARTAEINIVSKWQIKSLAFGPDHRIEMLEPEVLDGRMLKIWTDRAAETAQGLVY-DY 731
Db 687 PMDTEEAELVRQWENVAEALGPTHQVYSLSEVLDSMLVQW-QTLAQTAAEAKSCYWR 745
QY 732 TLKLSVDSVTSVAD---GTRALVEATLESACLSLIVHPENNATVRTYTRYEVFWMSK 788
Db 746 VLLHLEVLQAHIFEDGTIAGEAAEIEALLEBAELVDESQPK-NAKYTYTYKIRYILKKQE 804
QY 789 SG-WKITEGSV 798
Db 805 DGLWKFQCSDI 815

RESULT 12
Q55559 SYN3
ID Q55559 SYN3 PRELIMINARY; PRT; 714 AA.
AC Q55559;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE S110169 protein.
GN OrderedLocustNames-s110169;
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; BA000022; BAA10060.1; -; Genomic_DNA.
DR FIR; S76082; S76082.
DR HSSP; P08622; IBQZ.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0000917; P:barrier septum formation; IDA.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ; 1.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00076; DnaJ 2; 1.
KW Chaperone; Complete proteome.
SQ SEQUENCE 714 AA; 79424 MW; 51B52C16F405ED3E CRC64;
```

```
Query Match      10.6%; Score 429; DB 2; Length 714;
Best Local Similarity 25.1%; Pred. No. 1.7e-18;
Matches 194; Conservative 124; Mismatches 331; Indels 124; Gaps 30;

QY 86 IPIDFYQVLGAQTHFLTDGIRRAFEARVSPQPFQFSDALISRRQILQAACFTLSNPRS 145
DB 3 IPLDFYRILGIPQSGGETIEQAQVDRLLQLPRFSDAATVLENQALLATAYETLRPEK 62
QY 146 RREYNE---GLLDBE--EATVITDVPWVKP-----GALCVQEGGETEIVLVRVGEALLKE 196
DB 63 RQAYDQEWGAMDALGALPLATTPELECSPEQIEGALLILLDGEYELVVKYGEPLVD 122
QY 197 RLPKS--FKQDVVLWALAFDVSDDAMALDPDPFITGYEFVFEA-IKLQEEGASLAP 253
DB 123 PNPPAGGLPDYLLSVILAHWLSREWQQQ-----YEFATASLAKALARLOQDDNFP 176
QY 254 DLRAQIDETLEETPRVLELLGLP-LGDDVAAKRLNGLSGVRNLSVGG-GGASALVG 311
DB 177 ALBAEIRQELRYRPRYRILELLAKEGQEE--ORQGLALLQMWQDRGIEGKGEDYS 233
QY 312 GLTRE---KFMNEAFLMTAAEQVDLFVATPSNIPAESFEVYEVALALVAQAFTGKPKHL 368
DB 234 GLGNDDFLKTHQLRCHLTVAEQNALFL--PESQPSLVASYLAVHSLMAEGVKEQDPMA 291
QY 369 LQADKQFOQLQAKVWAMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVDCEKMWGLD 428
DB 292 IVEAKSLIIQLENCQ-----DLALEKVICELLIGQTEV--VLAAD 330
QY 429 SEDSQYRNPAIVFVLNENRDNDDLPGLKLETLWAGVVPFRDTPDKKFKLGDYY 488
DB 331 QGD-----PKIVA-GLSEKLTATGEDPLTAFYTFTEQWLEEBEIVPYFRDLSPETLSPKAYF 384
QY 489 DDPWLSYLERVEVQ-----GSP-LAAATWARIAGAEHVKASAM-QALQKVPFSPRYT 539
DB 385 NNFVQOYLEQLEPDSFTTNSPASPALLSTATSETPMWSSAALPDRPLTSTVPS--- 441
QY 540 DRNSAEKPVQOETVFSVDPVGNVGRGEPGVFAEAVRPSNFENNDYAIRA----- 592
DB 442 -RRGRSPRRSRDDVFP-----SADNSSGLAVT-TLSPAIAYDTHSLGTNGIGGDSTS 491
QY 593 -GYSESVDETTVE-MSVADMLKEASVK-----ILANGVAIGLISFSQK----- 635
DB 492 NGFSSNSAPESTSKHKSPPRRKRKRVTKPVRFGLFLLCLAGIVGGATALINRTGDP 551
QY 636 -----YFLKSSSSFORKDMVSSMESDVATIGSVRADDSALPRMDARTAEINVSQWK 688
DB 552 LLEDPLDVFL-----DQSEFIPDEATSNLILSQ-----PNFNQOQGVQVQWGLD 598
QY 689 IKSLAFGPDHRIEMLPEVLDGRMLKIWTDRAETAQLGLVYDYTLKLSDSVTS-ADG 747
DB 599 SKKLAFQNYDVGLAQSLVAPNLAAQQRG-AQRDQAQKYVHYEHKQLIAYQVNPQDP 657
QY 748 TRALVATELSEACSLDLVHPENNATDVR--TYTTRVEVFWKSGWKITBSVL 799
DB 658 NRATVTARVEEISQPTLGNQOQGSATKDDLTVRYQLVHQGVWKIDQIQVV 710

RESULT 13
Q9LJL2_ARATH PRELIMINARY; PRT; 841 AA.
ID Q9LJL2_ARATH
AC Q9LJL2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Arabidopsis thaliana genomic DNA, chromosome 3, p1 clone: MV111.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eutrosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
```

```
EX MEDLINE=20363099; PubMed=10907853;
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
RT and BAC clones.";
RN DNA Res. 7:217-221(2000).
RP NUCLEOTIDE SEQUENCE.
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP000419; BAB02958.1; -; Genomic DNA.
SQ SEQUENCE 841 AA; 93216 MW; 7DBEE72618EB8B97 CRC64;

Query Match      10.5%; Score 425; DB 2; Length 841;
Best Local Similarity 23.0%; Pred. No. 4e-18;
Matches 200; Conservative 122; Mismatches 308; Indels 240; Gaps 35;

QY 56 SSSSPATATTATLVSPPSIDPRHPVPIIDFYQVLGAQTHFLTDGIRRAFEARVSK 115
DB 81 SRTSSLAASTST-----IELPVTCTYQIGVSEQAQEKDEVKSVINLKKY 124
QY 116 PPOFGFSDDALISRRQILQAACFTLSNPRSREYNEGLLDDDEEATVITDVPWDKVPGALC 175
DB 125 DABEGYTWEAAAARQDILLMDVRDL---LFSEYAGNLKEKIAKPSLPIPAWLPGALC 181
QY 176 VLOEGGETEIVLVRVGEALLKERLPKSFQDVVLWALAFDVSDDAMALDPDPFITGYEF 235
DB 182 LLOEVQGEKLVLDIGRAALRNLDKPYTHDIFLSMALAECAIAKAAFEVNVKS--QGFEA 239
QY 236 VEEALKILOEB-GASSLAPDLRAQIDETLEITPRYVLELGLPLGDDDYAAKRLNGLSGV 294
DB 240 LARAQSFLLSKSVTLGKLA--LLTQTEESLEELAPPCTLDLGLPRTPENAEERRGAIAL 297
QY 295 RNILWVGSGGASALVGLTREK-----FMNEAFLRMTAAEQVDLF-----VATP 339
DB 298 RELRQ-----GLSVEASQIQDWPCFISQALISRLLAIEIVLLPWDDLIATRK 346
QY 340 SNTPAESFE-----VYEVALVAQAFIGKPKHLLQDADKQFQQLQAKVWAMEIP 390
DB 347 NKLSLESHNRQVITDFNCFYVWLVGHIAVGSF-----KQNETINKAKTICECLI 396
QY 391 AMLYDTRNNWEIDFGLERGLCALLIGKVDCE--RMWLGLDSE-----DSQY 434
DB 397 A-----SEGVDLKFEEAFCSFLLKQLSATGPTCWTIAQGSAAEAKLEKLQLESNDSAV 449
QY 435 RNPAIVFVFLNENRDNDDLPGLKLETLWAGVVPFRDTPK-----DKKF 482
DB 450 RNS-----ILKESRST-----ATPSLEAWMESVLANFPDTRGCSFSLANFFRAEKY 499
QY 483 KLGDIYDDPWLVS-----LERVEVVGSPALAAATMARIGAEHVKASAMQALQKFP 536
DB 500 PENKMGSPSIMNHKTNRPLSTTFVNSS-----QHL----- 532
QY 537 RYTRNSAEKPVQOETVFSVDPVGNVGRGEPGVFAEAVRPSNFENNDYAIRAGVSE 596
DB 533 -YTAVEQLTPTDLQSPVVSAK--NNDE-----TSASMPVQLKRN-----LGVHK 574
QY 597 SSV-DETTVMSVADMLKEASVKIL-----AAGVAIGLISLF-----SQ 634
DB 575 NKIWDW-----LSQSSLLGRVSVALLGCTVFFSLKLSGIRSGRLQSMPISVARPHSESD 631
QY 635 KYFLKSSSSFORKDMVSSMESDVATIGSVRA----- 665
DB 632 SFLWKTESGNFRKVLDSVNRNGI--VGNKIVLIDMLXMGHCEHPDALYLKSSGQSATSLS 689
QY 666 -DDSEALPR-MDARTAEINVSQWKIKSLAFGPDHRIEMLPEVLDGRM-----LK 713
DB 690 HSASELHKRPMDTEAEELRQWENKAEALGPTHQVYSLSEVLDESMLVQVESIFCLM 749
QY 714 INTDRAETAQLGLVY--DYTLKLSDSVTSVAD---GTRALVATELSEACSLDLVHPE 769
DB 750 QW-QTLTAQTAEAKSCYWRVFLVLLHLEQLAHIFEDGIAEAAEIBALLEEAELVDESQPK 808
```

```
QY 770 NNATDVRTYTRYEVFWSSKG-WKITEGVS 798
Db 809 -NAKYSTYKIRYILKKQDGLWKFQSDI 837

RESULT 14
Q8DKU7 SYNEL
ID Q8DKU7_SYNEL PRELIMINARY; PRT; 673 AA.
AC Q8DKU7;
DT 01-NAR-2003 (TrEMBLrel. 23, Created)
DT 01-OCT-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 25, Last annotation update)
DE Tlr0758 protein.
GN OrderedLocusNames=tlr0758;
OS Synecochoccus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecochoccus.
OX NCBI_TaxID=32046;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
CC -!- FUNCTION: Participates actively in the response to hyperosmotic
CC and heat shock by preventing the aggregation of stress-denatured
CC proteins and by disaggregating proteins, also in an autonomous,
CC dnaJ-independent fashion. Unfolded proteins bind initially to
CC dnaJ, upon interaction with the dnaJ-bound protein, dnaK
CC hydrolyzes its bound ATP, resulting in the formation of a stable
CC complex. GrpE releases ADP from dnaK; ATP binding to dnaK triggers
CC the release of the substrate protein, thus completing the reaction
CC cycle. Several rounds of ATP-dependent interactions between dnaJ,
CC dnaK and grpE are required for fully efficient folding. Also
CC involved, together with dnaK and grpE, in the DNA replication of
CC plasmids through activation of initiation proteins (By
CC similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- DOMAIN: The J domain is necessary and sufficient to stimulate dnaK
CC ATPase activity. Zinc center 1 plays an important role in the
CC autonomous, dnaK-independent chaperone activity of dnaJ. Zinc
CC center 2 is essential for interaction with dnaK and for dnaJ
CC activity (By similarity).
CC EMBL; BA000039; BAC08309.1; -; Genomic DNA.
CC GO; GO:0031072; F:heat shock protein binding; IEA.
CC DR GO; GO:0051082; F:unfolded protein binding; IEA.
CC DR GO; GO:006260; P:DNA replication; IEA.
CC DR GO; GO:006457; P:protein folding; IEA.
CC DR GO; GO:006986; P:response to unfolded protein; IEA.
CC DR InterPro; IPR001623; DnaJ_N.
CC DR Pfam; PF00226; DnaJ; 1.
CC DR SMART; SM00271; DnaJ; 1.
CC DR PROSITE; PS0076; DnaJ_2; 1.
CC KW Chaperone; Complete proteome; DNA replication; Heat shock.
CC SEQUENCE 673 AA; 7542 MW; 3427C6B46B83A CRC64;

Query Match
Best Local Similarity 25.4%; Pred. No. 3.8e-18;
Matches 191; Conservative 105; Mismatches 307; Indels 150; Gaps 28;

86 IPIDFYQLGAQTHFTDGIIRAFARVSKPPGFGSDALISRRQILQAACETLNPRS 145
|||:|||||
3 IPIDYYQLGLVPTQATPEQTEQAFDRLLQLPHTHQSPPTVATRRLLIEQAYAVLREPEQ 62
|||:|||||
146 BRYNE- -----GULDDEATVITDPWDKVPFGALCVLQEGETEIVLVRGEALL 194
|||:|||||
63 RDAYDRHCRVTDPDDLIAQLDPAATTPHIEISDEQLSGALLLYELGNVAQVNLGD AFL 122
```

```
QY 195 KERLPKSFQ-----DVIWMAFLDVSRDAMALDPDFITGYEFVFEALKLL 243
Db 123 KQDV---FERNRPTSPAAVADITLTVALAYLGRHEW--QRQSYESAASQLEAGLQVL 177
QY 244 QEEGASSLAPDLRAQIDETLEEITPRYVLELLGLPGDDVAAKRLNGLSGVNRILWSYGG 303
Db 178 QR---VNLPELOEQQTENLRPRYRILELLALPLSD--SANRQGIILLRQMLSERGG 232
QY 304 -GCASALVGLLTRE---KFMNEAFLRMTAAEQVDLFLVATPSNIPAESFEVVEYALALVAQ 359
Db 233 IEGRGDDRSGLTVEDFLKFILQLRSHLTVAEQQELF-----ERESRRPSAVATYLAHV 285
QY 360 AFIGKXPHILOD-----ADKFOOLOQAQVMMEIPAMLYDTRNNWIDFGLERGLCALL 414
Db 286 ALVARGVHQLPYSYICRAKDLLOL-----LPHQ-----DVIYELASCIALL 326
QY 415 IGKVDSCRMWGLGDSQVNRPAIVEFVLENSNRDNDLPGICKLLETWLAGVVRPRF 474
Db 327 LQOPTS-----ALALDHSQDQPTL-DFIRRHAG-EAGDRLPGLYYYTTQWLTEIYPAP 379
QY 475 RDTKDKKFKIGDYDDPMVLSYLERVEVVGQSPLAAATMARIGAHHKASAMQALQKVF 534
Db 380 RDLGETPVALEAYFADANVTYLEALSSEDSIAPEPPATTA----- 419
QY 535 PSRYDRNSAEPKDVQETVPSVDPVGNVGRDGEFGVFAEAVRPSNFENYDAIRAGV 594
Db 420 -----SALPEVIRPTV-AVPP-----PLSFTAETL-PLQDSRIGLOGLSASA 459
QY 595 SESSVDTEVMSVADMLKEASVK-----ILAAQVA-IGLISLFSQKFLKSS 642
Db 460 FTSATATGTSMPQSPRKRSPRNCARQRTWFMGAGVLVGLGAL--AKYVWPAPT 517
QY 643 SFORKDMVSMESDVATIGSVRADDSAL---PRMDARTAEINIVSKWQIKSLAFGPDHR 699
Db 518 AEAPPPVTPATPVATPTPTPQTTILAITLTPM-----ARDRLHTWQIIKAQALGRFFE 573
QY 700 IEMPEVLGRMLKIWTDRAAETAQLGLVYDYVYLLKLSVDSVTVS-ADGTRALVE----A 754
Db 574 VDKLTTLIAPELSRWRSAQGLKSGSYWVYVYTLKNLEKVEVRLQSRDRVLEAVNEDA 633
QY 755 TLESACLSDLVHPENNATDVRTYTRYEVFWS 787
Db 634 RFYEQGT-----RTDI-SYSDPYRVIYT 656

RESULT 15
O7V0H1_PROMP
ID Q7V0H1_PROMP PRELIMINARY; PRT; 702 AA.
AC Q7V0H1;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=PMW1287;
OS Prochlorococcus marinus subsp. pastoris (strain CCMP 1378 / MED4).
OC Bacteria; Cyanobacteria; Prochlorales; Prochlorococcaceae;
OX Prochlorococcus.
OX NCBI_TaxID=59919;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22825698; PubMed=12917642; DOI=10.1038/nature01947;
RA Rocap G., Larimer P.W., Lamerdin J.E., Malfatti S., Chain P.,
RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hess W.R.,
RA Johnson Z.I., Land M.L., Lindell D., Post A.F., Regalla W., Shah M.,
RA Shaw S.L., Stedlich C., Sullivan M.B., Ting C.S., Tolonen A.,
RA Webb E.A., Zinser E.R., Chisholm S.W.;
RT "Genome differentiation in two Prochlorococcus ecotypes reflects oceanic
RT niche differentiation.";
RL Nature 424:1042-1047(2003).
DR EMBL; BX572093; CAE19746.1; -; Genomic DNA.
DR GO; GO:0031072; F:heat shock protein binding; IEA.
DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR GO; GO:006457; P:protein folding; IEA.
```

Search completed: December 4, 2005, 12:50:06
Job time : 239 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 12:39:46 ; Search time 44 Seconds

(without alignments)
1751.582 Million cell updates/sec

Title: US-10-600-070B-2

Perfect score: 4063

Sequence: 1 MEALSHVGIGLSPFQLRLP.....YEVFWKSQWKITEGVSILAS 801

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------|---------------------|
| 1 | 497 | 12.2 | 798 | AD2144 | hypothetical prote |
| 2 | 429 | 10.6 | 714 | S7082 | hypothetical prote |
| 3 | 145 | 3.6 | 2297 | AB2494 | hypothetical prote |
| 4 | 143 | 3.5 | 2089 | A48757 | acetyl-CoA carboxy |
| 5 | 139.5 | 3.4 | 879 | JC4286 | DNA-directed DNA p |
| 6 | 138 | 3.4 | 2021 | A84771 | similar to ch-TOG |
| 7 | 131.5 | 3.2 | 1194 | DJBE28 | DNA-directed DNA p |
| 8 | 131.5 | 3.2 | 2111 | A70668 | mycocosate synth |
| 9 | 130.5 | 3.2 | 1026 | C97783 | cell surface antig |
| 10 | 129.5 | 3.2 | 2110 | B41110 | mycocosate synth |
| 11 | 129 | 3.2 | 1275 | T49362 | hypothetical prote |
| 12 | 128.5 | 3.2 | 836 | T30312 | pilin biosynthetic |
| 13 | 128.5 | 3.2 | 876 | S70368 | DNA polymerase I - |
| 14 | 127.5 | 3.1 | 2472 | E83594 | still frameshift p |
| 15 | 126.5 | 3.1 | 1018 | T30853 | antigenic heat-sta |
| 16 | 125.5 | 3.1 | 718 | A81122 | hypothetical prote |
| 17 | 125.5 | 3.1 | 4385 | T29042 | hypothetical prote |
| 18 | 123.5 | 3.0 | 673 | C83080 | probable chemotaxi |
| 19 | 123.5 | 3.0 | 848 | C70834 | probable endopepti |
| 20 | 123 | 3.0 | 3421 | WZBBE6 | 367K tegument prot |
| 21 | 122.5 | 3.0 | 899 | H87513 | hypothetical prote |
| 22 | 122.5 | 3.0 | 2756 | T30183 | hypothetical prote |
| 23 | 121 | 3.0 | 1381 | AF2010 | regulatory protein |
| 24 | 120 | 3.0 | 1252 | D72122 | RNA polymerase bet |
| 25 | 120 | 3.0 | 1262 | F81548 | DNA-directed RNA p |
| 26 | 120 | 3.0 | 3938 | T42761 | Bassoon protein - |
| 27 | 119.5 | 2.9 | 692 | G87415 | glycyl-tRNA synthet |
| 28 | 119.5 | 2.9 | 698 | A82593 | hypothetical prote |
| 29 | 119.5 | 2.9 | 831 | S26675 | DNA-directed DNA p |

| | | | | | | |
|----|-------|-----|------|---|--------|--------------------|
| 30 | 118.5 | 2.9 | 3942 | 2 | T42730 | Bassoon protein - |
| 31 | 117.5 | 2.9 | 1004 | 2 | G87323 | hypothetical prote |
| 32 | 117.5 | 2.9 | 4613 | 2 | T17409 | polyketide synthas |
| 33 | 116.5 | 2.9 | 880 | 2 | E69680 | DNA polymerase I p |
| 34 | 116.5 | 2.9 | 1074 | 2 | T01884 | hypothetical prote |
| 35 | 116.5 | 2.9 | 1644 | 2 | AC0823 | probable lipoprote |
| 36 | 116.5 | 2.9 | 3450 | 2 | T26963 | hypothetical prote |
| 37 | 116.5 | 2.9 | 3461 | 2 | T26964 | hypothetical prote |
| 38 | 115.5 | 2.8 | 693 | 2 | G82618 | pilus biogenesis p |
| 39 | 115.5 | 2.8 | 804 | 2 | T44606 | transducer protein |
| 40 | 115 | 2.8 | 1983 | 2 | AC1922 | two-component hybr |
| 41 | 114.5 | 2.8 | 758 | 2 | A98290 | exopolysaccharide |
| 42 | 114.5 | 2.8 | 758 | 2 | AB2994 | hypothetical prote |
| 43 | 114.5 | 2.8 | 1755 | 2 | F82618 | chemotaxis-related |
| 44 | 114.5 | 2.8 | 2333 | 1 | GNNY2F | genome polyprotein |
| 45 | 114 | 2.8 | 777 | 2 | T44597 | transducer protein |

ALIGNMENTS

RESULT 1

AD2144

hypothetical protein all2707 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004

C;Accession: AD2144

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriquchi,

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anab

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AD2144

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-798 <KUR>

A;Cross-references: UNIPROT:Q8YTL0; UNIPARC:UPI00000CE525; GB:BA000019; PIDN:BA074406.1;

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: all2707

C;Superfamily: Nostoc sp. hypothetical protein all2707; dnaJ amino-terminal homology

Query Match 12.2%; Score 497; DB 2; Length 798;

Best Local Similarity 24.6%; Pred. No. 3.9e-24;

Matches 207; Conservative 139; Mismatches 301; Indels 196; Gaps 33;

Qy 84 VPIDPYQVLGAQTHFLTDGIRAFARVSKPQFGSDDALISRRIQLQAACETLSNP 143

Db 11 VRPLDYRILGLPLAASDEQLRQAYSQVDRIVQLPRREYSQAIAIRKQLIEAYVVLSDP 70

Qy 144 RSREY-----NEGLLDDEEATVITDPWDKVGALCV 176

Db 71 KERSSVDQLYLAAHYDPDAAATTKVAENRGDSNNGHFDVOSLSI--EVSSEELGALLI 128

Qy 177 LQSGGTEIVLVRGEALL-----KERLPKSF-----KODVVLVMAALFDVSR 219

Db 129 LQELGEVELVLGRNVLGNQNGCTASTNGNHRPPEFLDSSERPDIILLTVALASLELGR 188

Qy 220 D-----AMALDPPDFITGVFEVEALKLQEGGASLAPDLRAQIDETUEEITPR 269

Db 189 EQWQOQHYEANAALSLE-----TQGE-----VLFSEG---IFPSVQAEIQADLYKLRPY 233

Qy 270 YVLELGLPLGDDYAAKRLNGLSVRNILMSVGG--GGASALVGGITRE---KFNWEAFRL 325

Db 234 RILELLALP--QEKTIERHQGLDLLQSLDDRGIDGTGNDQSGNLNIDDFLEFQQLRRH 291

Qy 326 MTAEOVDLFEVATPSNTPAESFEVYVALALVAQAFIGKPKPHLLQDADKQFOQLQAKVM 385

Db 292 LTVABQHLFDG--ESKKEPS--AVATYLAAYASIAAGFTQRPALIRHAKQILMRLSKRQ-- 347

Qy 386 AMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVDCECRMWLGDSDSQYRNPAIVEFVLE 445

149 YNE-----GLLDDEATVITVPMNDKVPGLACVLQEGGETEIVLVRGALLKER 197
:
:
:
312 FISVENITSLINLAALDDKDERTVFQOALEALPEKI-----ENGVN--IFSIIAPOLTEF 365
:
:
:
198 LPKSPKQDVVLWALAFLDVSRDAMALDPP-----DFTGVEFVEALKLLOEE-----G 247
:
:
:
366 EPQLEKTLSTLVNLRGQVQREALVAIAVLPKSEPIILQOALEIAIKIANESYDRIKA 425
:
:
:
248 ASSIAPDLRAQ-----IDTELEETIPRYVLELLGLPLGDDVAAKRLNGLSGVRNITLW 299
:
:
:
426 LAAIAPHLKETQCNEVLYLILERIENIWPFI-----SGVSETIV 465
:
:
:
300 SVGGGGASALVGLTRKFNWPAFLRTAAEQVDLFWATPSNIPAE--SFEVYVALALVA 358
:
:
:
466 FI-----ASYL--SKSQSKLLQKAF-----NIVQNLEFGSYDQAIALVA 502
:
:
:
359 QA--PIGK--KEPHLLQADKQFOOLOQ--AKUMAM-----EIPAMLYDTRNNWEI 402
:
:
:
503 LAPHLSKLEPOLLQOALKIAINIKEGASKVVALFAVPHLPQSKSOLLEKAFETIQTIEY 562
:
:
:
403 DFLGERLGCAL-----LIGKVDCKRMWGLDSEDISOYR--NPAP 439
:
:
:
563 DNARSRALVAVIPLHLSDFESHLLDKSLIEITKLVSTYDSAKA-QALVAVAPQLKFNPSL 621
:
:
:
440 VEFVLE-----NSNRDNDLPLGL-----KLETLWLAGVVPRFRDTKKKFLGSD 486
:
:
:
622 LQKALETAKKINSDEQKDEALAAIAYQLSESEPELLQYLKIAQNTVYVSHHHRIAIIP 681
:
:
:
487 YYDDPMWLSYLERVEVVGSPFLAAATWARIAGAHVKAASAMQ--ALQKVPFSRYTDRNSAE 545
:
:
:
682 YLREPOQTEFLQYFSATSSLASAAPCL---PEPKSEVLQELAKMLLDANYSDLLRAN 738
:
:
:
546 PKDQVETVFSVDPVGNVGRDGPFGVFIARAVPSENFTNDYAIRAGVSESSVDETTV-- 604
:
:
:
739 ELKIIIVPLSKT-----EQEKVEIEIAIKI--NDTFKGAELIAPVATHLSEFDQVKIF 790
:
:
:
605 EMSVADMLK---EASVKILAAAGVAIGLISLFSQKFLKSSSSSFQKMDVSMESDVAIG 661
:
:
:
791 FLBIEERLKAENDSSKAKALAVVIVPLSKSSPESLDKAFEIAENLQYQSCQPDDELTLA 850
:
:
:
662 S-VRAADSEALPRMDARTAEINIVSKWKIKSLA 693
:
:
:
851 THLKERECKILLEQALEKAKDIDSEYQQAQDFA 883

RESULT 4
A48757
C:Species: Cyclorella cryptica
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: A48757
R:Roessler, P.G.; Ohlrogge, J.B.
J. Biol. Chem. 268, 19254-19259, 1993
A:Title: Cloning and characterization of the gene that encodes acetyl-coenzyme A carboxylase
A:Reference number: A48757; PMID:93374903; PMID:8103514
A:Accession: A48757
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-2089 <R0E>
A:Cross-references: UNIPROT:Q39478; UNIPARC:UPI00000A5A26; GB:L20784; NID:g1065903; PIDN
A:Note: authors translated the codon GGC for residue 1834 as Ala
C:Genetics:
A:Introns: 25/1; 729/1
C:Superfamily: human acetyl-CoA carboxylase; biotin carboxylase homology; lipoyl/biotin
C:Keywords: biotin binding; ligase
F:99-603/Domain: biotin carboxylase homology <BCH>
F:731-803/Domain: lipoyl/biotin-binding homology <LPB>
F:770/Binding site: biotin (Lys) (covalent) #status predicted

Query Match 3.5%; Score 143; DB 1; Length 2089;
Best Local Similarity 20.0%; Pred. No. 1.7;
Matches 169; Conservative 113; Mismatches 289; Indels 272; Gaps 44;

QY 73 LPPSIDRPERHVPIDIFYQVLGAQTHPLTDGIRRAFEARVSKPP-----QGFSDAL 126

Db 197 LPNALDK-----LGKIFGTPGVMSVLGDKIAANILAQAKVPSIPWSGSGFGPDGP 250

QY 127 ISRQIILQACETILSNRPSREYNEGILD--DEEATVITDPWDKVPGALCVLQEGG--- 181

Db 251 LQ-----ADLTBEGTIPMEIFNKGVLVTGADEAVIVANKIGWEN--GIMIKASEGGGK 301

QY 182 -----ETEIVLR-----VCEALLKERLPKSFKODVYVVMALAFLDVSRDAMALDP 226

Db 302 GIRFVDEADLRNAPQVSNSEVGSPIFMQLCKNARHIEQIVG-----DQHGNAVALNG 357

QY 227 PDFITGYFVEEALKLQEGGASLAPDLRAQIDETLEEITPRVYVLELLGLPLGDDYAAK 286

Db 358 RDCSTQRRF-----QKIFEEGPPSIVP-----KETHEM-----ELAAQ 391

QY 287 RLNGLSGVRNILWSVGGGASALVGGITREKFWNEAFLEMTAAEOVDLFVA---TPSNIP 343

Db 392 RL-----TONIGYQ--GAGTVELYNAADNKFF---FLELNPRLQVEHPVTEGITGANLP 441

QY 344 AES-----PBYVEVALALVAQAFIGKPKPHLLQDADKQFOOLOQAKYMAEIPA--- 391

Db 442 ATQLQVANGIPFNIPDIRLYGREDAYGTP--IDPLQERYREL--DSHVIARITAEHP 498

QY 392 -----MLYDTRNN--EIDFGLERGLCALLIGKVDCECRMWLGDSQSY-- 434

Db 499 DEGFKPTSGSIERIKFGOSTPNVNGYPSVGANGGI-----HEFADSGFGH 542

QY 435 ---RNP-----AIVFVLENSNRDD--NDDLPLCLCKLET-----WLAGVVPF 472

Db 543 LFAKGNPREQARKALVLAKXEMEVRGDIRNSVYLVKLTETFAFKKNTIDTSLWDGII-- 600

QY 473 RFRDTKDKFKLDYDDPMVLVSYLEVERVVGQSPLAATAATMARIGAEHVKASAMQA--- 529

Db 601 -----KEKSVK-----EMPSHL-----VVGAAVFKAFEHVKVATEEVKESFRKGQVS 644

QY 530 -----LQKVPFSRYTDRNSAEPKDVQET----- 552

Db 645 TAGIPGINSFNIEVAYLDTKYPFHVERISPDVYRFTLDGNTIDVEVTQTAEGALLATFGG 704

QY 553 -----VFSVD--PVGNVNGRGE-----PGVFIAEAVRPS-----ENFETNDYATRAGVSES 597

Db 705 ETRHIFGMDEPLGLRLSLDGTAVLMPITFDPSSELRDTDVTGKVRYLQDNGATVEAG--- 760

QY 598 SVDETTVMASVADMLKEASVKILAAAGVAIGLISLFSQKFLKSSSFQKQDMVSSME-SD 656

Db 761 ---QPYVEVEAMKMI--MPIKATESG-----KITHNLSAGSVISAGDLLASLELKD 806

QY 657 VATIGSV-----RADDSREALPRMDARTA--ENIVSKMQKIKSLAFGPDHRIEMLPVELDGR 710

Db 807 PSRVKKIETFSGLDIMEKVDLEPQKAVNVLS-----GFNLD-----PEAVAQ 852

QY 711 MLKIWTDRAETRAQLGLVYD--YTLCLKLVDSVTVTSADGTALVEATLESACLSDLVHPE 769

Db 853 AIDSATDSSAAADLLVQVLDEFYRVESQFDGV-IADDVVVRLTKANTE-----TLDVVVISE 907

QY 770 NNA 772

Db 908 NLA 910

RESULT 5

JC4286

DNA-directed DNA polymerase (EC 2.7.7.7) - Bacillus stearothermophilus

N;Alternate names: DNA polymerase I

C;Species: Bacillus stearothermophilus

C;Date: 14-Nov-1995 #sequence_revision 08-Feb-1996 #text_change 17-Mar-1999

C;Accession: JC4286

R;Phang, S.M.; Teo, C.Y.; Lo, E.; Wong, V.W.T.

Gene 163, 65-68, 1995

A;Title: Cloning and complete sequence of the DNA polymerase-encoding gene (Bst)

A;Reference number: JC4286; MUID:96001245; PMID:7557480

A;Accession: JC4286

A:Molecule type: DNA
A:Residues: 1-879 <PRA>
A:Cross-references: UNIPARC:UPI00001757A1; GB:U23149
A:Note: The authors translated the codon GCG for residue 15 as Arg
C:Comment: It has three enzymatic functions: 5'-3' exonuclease activity, DNA synthe: is a
C:Genetics:
A:Gene: Batpol1
C:Superfamily: DNA-directed DNA polymerase I
C:Keywords: nucleotidyltransferase

Query Match 3.4%; Score 139.5; DB 2; Length 879;
Best Local Similarity 22.0%; Pred. No. 0.71;
Matches 175; Conservative 132; Mismatches 315; Indels 173; Gaps 44;

Qy 99 HFLTDGIRRAFAEARKVPQFGSDDALISRRQLQAACETLSNPRSRREY-NEGLLDE 157
Db 108 HYEADDIIGTWAARAERE---GFAVKVISGRDLTQ-----LASPQVTVEITKKGITDIE 159
Qy 158 EATVITDV-----PNDKVPG-----ALCVLOEGGETEIVLRVG 190
Db 160 SYTPETVEKYGLTPEQIVDLKGLMGDKSNIPGPGIGKKTAVKLLKQFGTVENVLASI 219
Qy 191 EALLKERLPKSFQK--DVLVMAFAFLDVSQDA---MALDPPDPFITGVZFVEEALKLQ 245
Db 220 DEIKGEKLENLQRYRDAL-LSQLAAICEDAPVEITLD--DIVYGEDEBKVVAFQ 276
Qy 246 EGASSLAPDLRAQIDE-----TLBEITPRYVLELGL---PLGDDYAAKRLNG- 290
Db 277 LGFQSFLLKMAVQTDEGEKPLAGMDFADSVTDEMLADKAALVVEVGVNYHHAPITGI 336
Qy 291 -LSGVRNLM-----SVGGGASALVGLTRFKFM-----NEAFLEMTAAEOV-D 333
Db 337 ALANERGRFLPETAVADPKFLAWLGDETKKTKTFDSKRAAVALNGKIELAGVGVVFD 396
Qy 334 LFVATPSNIPAES-----PEVVEVALVAQAFIGK-KPHLLQDADKQFOOL--OOAK 383
Db 397 LLLAAYLLDPAQAAGDVAANAKMOYEAVRSDAEVYGGKAKRTPDPTAEQLVRKAAA 456
Qy 384 VMAMEIPAMLYDTRNWE-IDFGLERGLCALLIG-----KVDECRM-WLGLD-SEDSQY 434
Db 457 IWALEEPLMBELRNEQDRLLTELEHALAGILANMEFTGVKVDTKRLQONGAEITEQLQ- 515
Qy 435 RNPAIVEFVLENSRNDNDLLPLGLKLETLWLAGVFPFRDTRDKKFKLGDYDDPMVL 494
Db 516 ---AVERRIYELAGQEFNINSP---KQLGT-----VLFDKLQPLVKTKTKTG-YSTSADVL 564
Qy 495 SYL-ERVEVVOGSPLAATAATMARIGAEHVKASAMQALQKVPFSRYTDRNSAEPKDVOET- 552
Db 565 EKLAPHHEIVE--HILHYRQLGKLOSTYIE-GLLUKVHPVTVGKVHTWPNQA-----LTQTG 617
Qy 553 -VFSVDPVGNNGRDBGPGVPIABAVRPSB-----NFETNDYAIRAGVSSESSVDETT 603
Db 618 RLSSVEPLQNIPIRLSEGRKIROAFVPSBPDWLI PAADYSQIELRVLAHIAE---DDNL 674
Qy 604 VE-----MSVADMLKE---ASVKILAAAGVAIGLISLFSOKYFLKSSSSPQK 647
Db 675 IEAPRRWLDIHTKTAMDIFHVSSEEDVTANMRQAKAVNFGIVYGISD-YGLAQNLIATR 733
Qy 648 DMVSSMESDVATIGSVRADSEALPRMDARTAEINVSQWKIKSLAAGPDHRIEMLEVL 707
Db 734 EAABEFIRYFASFPGVK-----QYMDNIVQAE-AKQGYVTTLLHRRRYLPDIT 780
Qy 708 DGRM-LKIWTDRAA-ETAQGLGVYDVTLLKLSVD-SVTVSADGTRA-----LVSEA 754
Db 781 SRNFNVTFAERTAMNTPIQGSAAD-IKKAMIDLVSVRERLQARLLQLQGHDELILEA 839
Qy 755 TLEESACLSLWHP 769
Db 840 PKEBIGRLCRLV-PE 853

RESULT 6
A84771

similar to ch-TOG protein from Homo sapiens [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: A84771
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84771
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2021 <STO>
A:Cross-references: UNIPROT:Q9ZQNG; UNIPARC:UPI000017A6E6; GB:AE002093; NID:g4263790; PIR
C:Genetics:
A:Gene: At2g35630
A:Map position: 2

Query Match 3.4%; Score 138; DB 2; Length 2021;
Best Local Similarity 20.2%; Pred. No. 3.3;
Matches 179; Conservative 122; Mismatches 315; Indels 272; Gaps 43;

Qy 39 ASKWADRLSDNFNTSDSSSSSPATATTTATLVSLSIDRPERHVPIDFYQVILGAQT 98
Db 297 ATWSEKREAVELTKLASTKKIAGDSEICRTKKLI--TDVNLAVAVEAIAQIGN-- 352
Qy 99 HFLTGTGIRRAFAEARKVPQFGSDDALISRRQLQAACETLSNPRSRREYNEGLLDEE 158
Db 353 --LACGLRTHFSA-----SSRFMLVPLEKL-----KEKKQSVTDPLT 388
Qy 159 ATVITDVPMDKPGALCVLOEGGETEIVLRVGEALLKERLPKSPKQDVVLVMA-----LA 213
Db 389 QTLQT-----MYKAG-----CLNLVDVIEGKVNKTAVKNKVPPLVRSSTLTWLT 431
Qy 214 F-LDVSERDAMALDPPDPFITGVZFVEEALKLQEGASSLAPDLR-----AQID----- 260
Db 432 FCLETSNKALILK-----AHKEYVPLCMECLNDG-----TPDVRDAAAFSAALAAKSVGM 481
Qy 261 -----ETLEBITPRYVLELGLPLGDDYAAKRLNGLSGVRNILWSVGGGASALVGL 313
Db 482 RPLERSLEKLDVKKKLSMIAGSGGDOA-----GTSSV-TVQSSVSGSTATGCLHNRY 535
Qy 314 TREKFM---NEAFLEMTAAE-----QVDLFVAT-----PSNI 342
Db 536 TTSKFMQNSDASFVRKSAASMLSGKRPAPSAQASKVGTGKPGGKKDGSVRNEGSKSVE 595
Qy 343 PAESPEVVEVALVAQAFIGK--KPHLLQDADKQF--QOLQAKVMAEIPAMLYDTRN 398
Db 596 PPEDEVEPAEMGLEIENR-LGSLVKPETSQLSKSVWKERLEATLAKKEIEGL----- 648
Qy 399 NWEIDFGLE---RGLCALLIGKVDCEKRMVLGLDSEDSQYRNPAIVEFVLEN---SNRDD- 451
Db 649 -QELDKSVELLRLLCAV-----PGWKEKVQL-----FSFENYFNSLPDF 689
Qy 452 -----NDDLPGCLKLETLWAGVFPFRDTRDKKFKLGDYDDPMVL 494
Db 690 SINSLGTSERVADIKTRASAMKCLTAPCEAVGPGFVPERL-----FKIMKEHKNPKVL 742
Qy 495 S-----YLERVEVVOGSPLAATAATMARIGAEH----- 521
Db 743 SEGILLMWVSVDVFGVSLLLKDLIDFCQDVGLQSSTAATRNATIKLLGALHKFVGPDIK 802
Qy 522 -----VKASAMQALQKVPFSRYTDRNSAEPKDVOETVFSVDPVGNVGRDCEPGVFAEA 576
Db 803 GFLNDVKKPALLSADTET-EKNPPEGTAAAPRVVKT--SVSTSSCGGLDLPREDISTK 859
Qy 577 VRPS--ENFETNDYAIRAGVSSESSVDETTVMESVADMLKEASVKILAAAGVAIGLISLFSQ 634
Db 860 ITPNLLKGFESPDWKML-----ESIE-----AVNKILEEANKRIQPTGTG----ELFGG 905
Qy 635 KYFLKSSSSFORQDMVSSMESDVATIGSVRADSEALPRMDARTAEINIV-----SKWQK 688

```
Db 906 ---LRGLLDNKNLVM---QTLTTIGVAAAMGPAVEKASKGILSDVLKLCIGDNKKHMR 959
Qy 689 IKSLA-----FGPHRRIMLEPEVL-----DGRM-----LKIWTDRAAETAQLGLVVDYT 732
Db 960 ECTLAALDLMLGAVHLMDPIYIIALTIDGKMGAEGRKDLFDML-----TKQLTGLSDFV 1014
Qy 733 ----LLKLSVDSVTVSADGTRALVEATLESACLSDLVHPENNATDVR 776
Db 1015 DAHLLKPASTAMTKSADVRKAAEGCISEILRVSGQEMIEKNLKDQ 1062

RESULT 7
DJBE28
DNA-directed DNA polymerase (EC 2.7.7.7) - human herpesvirus 3
C:Species: human herpesvirus 3, varicella-zoster virus
C>Date: 30-Sep-1998 #sequence_revision 30-Sep-1988 #text_change 09-Jul-2004
C:Accession: B27214
R: Davison, A.J.; Scott, J.E.
J. Gen. Virol. 67, 1759-1816, 1986
A:Title: The complete DNA sequence of varicella-zoster virus.
A:Reference number: A27345; MUID:86306657; PMID:3018124
A:Accession: B27214
A:Molecule type: DNA
A:Residues: 1-1194 <DAV>
A:Cross-references: UNIPROT:P09252; UNIPARC:UPI0000129829; EMBL:X04370; NID:959989; PIDN
C:Genetics:
C:Superfamily: DNA polymerase
C:Keywords: DNA binding; DNA replication; nucleotidyltransferase

Query Match 3.2%; Score 131.5; DB 1; Length 1194;
Best Local Similarity 19.0%; Pred. No. 3.8;
Matches 168; Conservative 122; Mismatches 297; Indels 297; Gaps 44;

Qy 48 SDNFETSDSSS---SSPATATTATVLSLPPSIDRPERHV--PIPIDFYOVLCQAQ---TH 99
Db 205 NDATLNGDKNAFTGTSKSPSPFRVEV---IERIDVYTDQPCAFYRVYSPSSKFTN 261

Qy 100 FLTDGIR---RAFEARVSKPPQFGFSDALIS-----RRQI--- 132
Db 262 YLCDNFHPELKKYEGRVDAITRLMDNPGFVSGWYQLKPGVDGVRVRVPASRQLTSD 321

Qy 133 LQAAQ-----ETLSNPRSRREYNGLLDDR-----EATVTDVPVDPKVPQALC 175
Db 322 VEIDCMSNLQAI PNDDSPWDYKLLCFDIECKSGSNELAFPDATHLEDL---VQIQC 377

Qy 176 VL---QEGGETIVLRVGEALKERLPKSFQ-----DVLVWALAFLDVSRD 220
Db 378 LLYSIPRQSLHILLFSLGCDLPQRYVQEMKDAGLPEPTVLFFDGEFELLIAFMTLVKQ 437

Qy 221 AMALDPPDFTGYEFVEEALKLQEBGASSLAPDLRAQIDETLEEITPRVLELLGLPLG 280
Db 438 Y---APEFATGNIYNFDWAFI-----MEKLSIYSLKL----- 468

Qy 281 DYAAKRLNGSLGVRNLMVSGGGASALVGGITREKFMNEAFRLMTAAEQVDLFPATPS 340
Db 469 DVGSGINRGLFKI---NDVGKSG-----FQRRSKVKINGLISLDWYALATE 512

Qy 341 NIPAESFEVVEVA-----LALVAQAFIG 363
Db 513 KLSLSYKLDSDVAREALNESKRDLPIKYDIPGYVYASGNTGRITIGEYCIODSALVGLKFFK 572

Qy 364 KPHLLQDADKQFQQLQQAQKWAWEIPAMLYDTRNNWEIDFGLERGLCALLICKVDECRM 423
Db 573 YLPHLELSA-----VARLARITUTKAIYD-----GQVRIYTCLLGLASS--- 612

Qy 424 WLGLDSEDSOYRNPAIVEF-----VLENSNRDNDLPLGCKLLETWLAGV-VFPR 473
Db 613 -RGFILLPDGYY--PATFEYKDVIPDVGDEEENDESDSPGTSNGRNVGYKGVARVD- 668

Qy 474 FRDTKDKKFKLGDYDDPWLVSLEVRV-EVVGSPPLAAATATWARI GAHVKASAMQALQK 532
Db 669 -PDT-----GFYIDPVVVLDPASLYPSIIQAHNL--CFTTILTNPETVK-----R 710
```

```
Qy 533 VFPGRYTDNRNABPKDVOETVFSVDP-----VGNVNRDGBEGCVFAIE--AVRPSNPFT 585
Db 711 LNPSDY-----ATFTVGGKRLFFVRSNV-RESLLGLVLLKOWLAMRK----- 750

Qy 586 NDYAIRAGVSSSDVDETTVEMSVADMKEASVKILA-----AGVAIGLIS----- 630
Db 751 ---AIRARIPGSSDE-----AVLLDKQAIAKVVCSNVYGTGVAQGFPLCLYVAATVT 802

Qy 631 -----LFSQKYLKSSSFQR-----KDMVSSMESDVA-TIGSVRRADDSSEALPRMDAR 677
Db 803 TIGRQMLLSTRDIHNNWAAFERFITAPDIESSVLSQKAYEVKVIYDGTDSVFIRFKGV 862

Qy 678 TAENIVSKWQK-----IKSLAFGPDPHRIEMLEPEVL DGRMLKIWTDRAAETAQLGLVVDYT 733
Db 863 SVEGIATIGERWAHIIISTALFCPPIKLECEKTFI--KLLLI-----TKKKYIGVIYGGKV 915

Qy 734 LKLSVDSVTVS-----ADGTRALVEATL-----EESACL 763
Db 916 LMKGVDLVRKNKCOFINDYARKLVLELLYDDTVSRAAAEASCVS 959

RESULT 8
A70668
mycocerosate synthase (EC 2.3.1.111) - Mycobacterium tuberculosis
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C:Accession: A70668
R: Cole, S.T.; Broech, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70668
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2111 <COL>
A:Cross-references: UNIPROT:P66291; UNIPARC:UPI0000036BA0; GB:Z83858; GB:AL123456; NID:959989;
A:Experimental source: strain H37Rv
C:Genetics:
C:Gene: mas
C:Superfamily: mycocerosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I hom
name homology; [acyl-carrier-protein] S-malonyltransferase homology
C:Keywords: acyltransferase; carrier protein; coenzyme A; phosphotransferase
F:27-426/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F:534-815/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F:1444-1733/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:1765-1945/Domain: short-chain alcohol dehydrogenase homology <SADH>
F:2016-2096/Domain: acyl carrier protein homology <ACP1>

Query Match 3.2%; Score 131.5; DB 2; Length 2111;
Best Local Similarity 19.9%; Pred. No. 9.4;
Matches 151; Conservative 82; Mismatches 237; Indels 289; Gaps 33;

Qy 151 EGLLDDEEAATVTDVP---WDK-----VPCALCVLOEGGETEIVLRVGEALLKERL 198
Db 29 ESLLRGDD---LVTEIPDRWDADDYDPEQVPGVR-SVSRWGG----- 68

Qy 199 PKSFQKQDVILVMAFLDVSRDAMALDPPDFI---TGVEFVEEALKLQEBGASSLAPDL 255
Db 69 ---FLDDVAGDAEFFGISEREATSIDPQQRLLLETSWEAIEHA----- 109

Qy 256 RAQIDETLEEITPRVLELGLPLGDDYAAKRLNGLSGVRNLMVSGGGASALVGGITR 315
Db 110 -----GLDPASL-----AGSSTAVFTGLTH 129

Qy 316 EKFMNEAFLRMTAAEQVDLFPATPSNI-----PAESFEVY----- 350
Db 130 EDYL---VLTTTAGGLASPVVTVTGLNNSVASGRIHTLGLHGPANTFDTCSSGLMAVHL 186

Qy 351 -----EVALALVAQAFIGKPKPHLL-----QDADKQFQLOQAK 383
```

```
Db 187 ACRSLHDEADLALAGGCAVLEPHASVAASAGCMLSSTGCRCHSFADADGFFVRSEGCAM 246
Qy 384 VNAMIEPAMLYDTRNNWEIDFGLRGCLCALLIGKVDCEKRMWLGDSQYRNPAIVEFV 443
Db 247 VLLKELPDALRDGNR-----IPAVRGATATNQDGRTE-----LTPSEDAQV---AVYRAA 295
Qy 444 LENSRRDNDLPLGLCKLETWLAGVVPFRDTRDKKFKLGDDYDDPMVLSYLERVEV 503
Db 296 LAAAG-----VQPETGVV-----EAHGTGTPIGDPIEYRSLARV-YG 332
Qy 504 QCSPLAAATWARIAGHVKASAMQALQKV-----PSRYTDRNSAEPKDVQETVFS 555
Db 333 AGTPCALGSAKSNMGHSTASATVGLIKAILSLRHGVVPLLHFNRLPDLSDV-----386
Qy 556 VDPGVNNGVRDEPGPVFAEAVRPSSENFETNDYA-IRAGVSESSVDDETTVMSVADMLKE 614
Db 387 -----ETGLFVQAVTPWPN--GNDHTPKRVAVSFGMSGTNVHAIVEAPAE 432
Qy 615 ASVKILAGVA-IGLISLFSQKYLKSSSFORKOMVSMESDVATIGSVRADDSEALPR 673
Db 433 ASAPESPQDAEAVG-----PRLFMLSSTS-----SDAL-R 461
Qy 674 MDARTAEINIVSKWK-----IKSLAFGPDHR-----IEMLPVLDGRML 712
Db 462 QTARQLATVVEHQDCAASDLAYTLARGRAHRPVRTAVVAANLPVLVGLREVADGDAL 521
Qy 713 -----KIW--TDRAAETAOLGL-----VYDYLKLSVDSVTVSADGTRALV 752
Db 522 YDAANVGHGDRGPVWVPSGGSQWAAMGTOLLASEPVFAATIAKL---EPVIAESGFSVT 578
Qy 753 EA-TLEESACLSDLVHPENNATDVRTYTYRYEVFWSKSG 790
Db 579 EAITAQOVTGIDKQPAVFAVQVALAATMEQTYGVRRPG 617

RESULT 9
C97783
cell surface antigen [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: C97783
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: C97783
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1026 <KUR>
A:Cross-references: UNIPARC:UPI0000164BP3; GB:AE006914; PIDN:AAL03205.1; PID:g15619.56;
C:Genetics:
A:Gene: sca4

Query Match 3.2%; Score 130.5; DB 2; Length 1026;
Best Local Similarity 18.8%; Pred. No. 3.5;
Matches 124; Conservative 109; Mismatches 277; Indels 151; Gaps 25;

Qy 115 KPQGFSDALISRRQILQAAEILSNPRSRREYNEGILLDBEATVITDPWQKV----170
Db 378 EEPYQKQVPPITPTNQPLQ--ETSQMPQS--QVNPNNLN--TATALSGSQMDLLNYVN 432
Qy 171 PGALCVLQGGGTEIVLRVGEALLKERLPKSPKQDVVLVMA--LAFIDVSRDAMALDPPD 228
Db 433 AGLTAKIDSNKLDLKEATAILNE-----KSDIAEQANTIALAENTVNNKLNKLPDA 487
Qy 229 FITGYEFVEEALKLQEGGASSLAPDLRAQI-----DETLEIITPRYVLEL---LGLPLG 280
Db 488 KTVGNVAVLETKINDQNTENLEKSKMLEATVAIVLNSLENLPKQKQMLEKAVDVGLSLK 547
Qy 281 DYAAKRLNGLSGVRNILMSVGGGASALVGGUTREKFNWEPFLRWTAAEQVDLFWATPS 340
Db 548 DD--ASRAAIDGIKDV-----IKSNLSPEDKMLIAVGD 580
```

```
Qy 341 NTPAESFEVTEVALVAQAFIGKKPHLLQADAKQFOQLQOAKVMA-----MEIPAM 392
Db 581 KVNVSLSNAEKOKLGLSVLKGVGEAVQLSPAQOQLMQOHLKYKIMAEQTKDKTKKVNDI 640
Qy 393 LYDTRNWE-----IDFGLRGCLCALLIGKVDCEKRMWLGDSQYRNPAI 439
Db 641 LFPPLSNTLTKTNNIOAITSNVLDGPATAVKGEIIQAITNTIAGSSLEAOD-----KAAI 696
Qy 440 VEVPLEN--SNRDDNDLPGCLCKLETWLAGVVPFRDTRDKKFKLGDDYDDPMVLSYL 497
Db 697 IKGVETIATHSDTSLSPNKALIMASAEKGI-----ESQTNLPD--RELMTKGLV 746
Qy 498 ERVEVVGSPPLAAATWARIAGHVKASAMQALQKV-----FPRSRYTDRN 542
Db 747 DGIYEGKGPEITKAVSSGIDNSINDSEKEALKKAKDAASEAALDRDTONLTGFKGQN 806
Qy 543 SASPK-----DVQETVFSVDPVGNNGVGRDGPGEVIAE-----AVRP 579
Db 807 IEHKKPHDDIYNKAREVINAVNPVIEALEKSKPVSABERIVQETSILNNISKLAVEK 866
Qy 580 SENFETNDYAIRAGVSESSVDDETTVMSVADMLKEASVKILAGVAIGLISLFSQKYLK 639
Db 867 VNNF-----RAWLSPNGNLKLEK-----KEBAIKKVE-----LVKAFGTKSTE 908
Qy 640 SSSSFORKOMVS-----SMESDVATIGSVRADD--SEALPRMDARTAE-NIVSKWKQIKSL 692
Db 909 EQQSFIKTNLIDDKLTKSKEVRLQTLKLEQKRSEAIENPSVKTEDVVRVVSQKSLKPI 968
Qy 693 A 693
Db 969 S 969

RESULT 10
B44110
myceroosate synthase (EC 2.3.1.111) - Mycobacterium bovis
C:Species: Mycobacterium bovis
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 09-Jul-2004
C:Accession: B44110
R:Nathur, M.; Kolattukudy, P.E.
J. Biol. Chem. 267, 19388-19395, 1992
A:Title: Molecular cloning and sequencing of the gene for myceroosic acid synthase, a n
-Guerin.
A:Reference number: A44110; MUID:92406887; PMID:1527058
A:Accession: B44110
A:Molecule type: DNA: protein
A:Residues: 1-2110 <MAT>
A:Cross-references: UNIPROT:Q02251; UNIPARC:UPI000012ED5D; GB:M95808; NID:g149977; PIDN:
A>Note: sequence extracted from NCBI backbone (NCBIN:113897, NCBIPI:113899)
A>Note: the source is designated as Mycobacterium tuberculosis var. bovis Bacillus Calmet
C:Superfamily: myceroosic acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I hon
nase homology; [acyl-carrier-protein] S-malonyltransferase homology
C:Keywords: acyltransferase; carrier protein; coenzyme A; phosphotransferase
F;27-426/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F;534-815/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F;1443-1732/Domain: long-chain alcohol dehydrogenase homology <LADH>
F;1764-1944/Domain: short-chain alcohol dehydrogenase homology <SADH>
F;2015-2095/Domain: acyl carrier protein homology <ACPI>

Query Match 3.2%; Score 129.5; DB 2; Length 2110;
Best Local Similarity 19.9%; Pred. No. 13;
Matches 151; Conservative 82; Mismatches 237; Indels 289; Gaps 33;

Qy 151 EGLDDDEATVITDVP---WDK-----VPGALCVLQGGGTEIVLRVGEALLKERL 198
Db 29 ESILRGDD--LVTEIIPDRWDADDYDPEPGVGR-SVSRWGG-----68
Qy 199 PKSFQKDVVLVMAALFLDVSRDAMALDPPDFI---TCGYEFVEEALKLQEGGASSLAPDL 255
Db 69 ---FLDDVAGDFAEFFGISEREATSIDPQORLLLETSWEAIEHA-----109
Qy 256 RAQIDETLEBITPRYVLELIGLPLGDDYAAKRLNGLSGVRNILMSVGGGASALVGGILTR 315
```

```
Db 110 -----GLDPASL-----AGSTAVFTGLTH 129
Qy 316 EKFWNEAFRLMTAAQVDLFWATPSNI-----PAESFEVY----- 350
Db 130 EDVL---VLATTTAGGLASPYVVTGLNNSVASGRIATHTLGLHGPAMTPTDACSGLMAVHL 186
Qy 351 -----EVALALVAQAFIGKPKHL-----ODADKQFOOLOQAK 383
Db 187 ACRSLHGDRLDALAGGCAVLELHACVAASQAQMLSGTGRCHSFDADAGFVRSEGCAM 246
Qy 384 VMAMEIPAMLYDTRNNNEIDFGLERGLCALLIGKVDRCMWLGLSDSDSOYRNPAIVEFV 443
Db 247 VLLKXLPDALRDGNN---IFAVVRGTATNQDGTET---LTPSDDAQV---AVTRAA 295
Qy 444 LENSNRDNDLPLGLCKLKTWLAGVYFPRFDTKKFKLGDYDDPMVLSYLERVEV 503
Db 296 LAAAG-----VQPETGVV-----EAHGTGTPIGDPFIEYSLARV-YG 332
Qy 504 QGSPAAAAATMARIGAHHVKAAMQALQKVF-----PSRYTDRNSAEPKQVQETVFS 555
Db 333 AGTFCALGSAKSNHGHSASAGTVGLIKAILSLRHGVVYVPLLHFNRLPDELSDV----- 386
Qy 556 VDPGVNNVGRDGPVFIARAVPSENFETNDYA-IRAGYSESSVDVETTVMSVADMLKE 614
Db 387 -----ETGLFVQAVTPWPN--GNDHTPKRVAVSSFGMSGTGNVHAIVEEAPAE 432
Qy 615 ASYKILAAAGVA-IGLISLFSQKFLKSSSSSFQKDMVSSMESDVATIGSVRADDSALPR 673
Db 433 ASAPESPGDAEVG-----PRUFMLSTS-----SDAL-R 461
Qy 674 MDARTAEIVSKWKQ-----TKSLAFGPDHR-----IEMLPVLDGRML 712
Db 462 QTAARQATWVEEHQDCVAASDLATYLARGRRHPRTAVVAANLPELVEGLREVADGDAL 521
Qy 713 -----KIW--TDRAAEFAQIGL-----VYDYTLKLSVDSVTVSADGTFALV 752
Db 522 YDAVGHGDRGPMVWFSQGQSOAAAGTQLLASEPFAATIAKL---EPVIAAESGFSVT 578
Qy 753 EA-TLESACLSDLVHPENNATDVRTVTRYEVFWSKSG 790
Db 579 EAITAQQTVTGIDKQVPAFVQVVALAANTWEQTYGVYRPG 617

RESULT 11
T49362
hypothetical protein B1D1.160 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C:Accession: T49362
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A:Reference number: Z25022
A:Accession: T49362
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1275 <SCH>
A:Cross-references: UNIPROT:Q9P6A4; UNIPARC:UPI000006B2D3; EMBL:AL355927; GSPDB:GN01116;
A:Experimental source: BAC clone B1D1; strain OR74A
C:Genetics:
A:Gene: NCSP:B1D1.160
A:Map position: 6
A:Introns: 24/3

Query Match 3.2%; Score 129; DB 2; Length 1275;
Best Local Similarity 20.9%; Pred. No. 6.2;
Matches 114; Conservative 84; Mismatches 210; Indels 138; Gaps 26;

Qy 325 RMTAAEQVDLFWATPSNIPAESFEVVEVALVAQAQFIGKKPHL----- 368
Db 39 RNTPATPTAYFAALLSLNNENIATPVVLLDVTVPF-APKPILOAKFTQITILAPVLS 97
Qy 369 LQADKQFOOLOQAKVMAMEIPAMLYDTRNNWEI---DFGLERGLCALLIGKVDCEMWL 425
```

```
Db 98 LPDAD---APLMRASIGSLE-SLLLAQDATQWELTAQIGPRRAVAGLLA-----L 144
Qy 426 GLDSEDS-QYRNPAIVEFVLENSNRDNDLPLGLCKLJLKTWLAGVYVFPFRDTKD----- 479
Db 145 ALDPRPKVKEQAQRALRNILKNPPSPSLDHPAAPMCAETAMQSL-----RLVADKAAQA 199
Qy 480 -KKFKLGDYDDPMVLSYLERVEVQ-----GSPAAAAATM-----ARIGAHHVKAAM 527
Db 200 RKEKTTDSTHDPDLIHALQLVKAVASGGGWPFSKSIESLCELLSIARTGNEHMSMAVF 259
Qy 528 QALQKVPSPRYTDR-NSAEPKDVQETVFSVDPVGNV-----VGRDGPVGFIAEA 576
Db 260 EIFEMFEQMADEVASAKLPRLEIIRRLPAPNDTQLLPWIAILSRAYD-----VAQ 315
Qy 577 VRPSENFEE--TNDYAIRAGVSES-----SVDETTVMSVADMLKEASV-----KIL 620
Db 316 ISPARTQELVDPTFLVAGYLESQHKNIRVSAECLVSLANCVPKEAILEPSIDPKVI 375
Qy 621 AAGVAI--GLISLFSQKFLK-----SSSFQKDMVSSMESDVATIGSVRADDS 669
Db 376 QQLVKVVEGLLTVOYQAAMMETFNVLGAMFADFARHQANPYLLSV---VKSIGEIRGNSF 432
Qy 670 ALPRMDARTAEIVSKWKIKSLAFGPDHRIEMLP-----EVLDCRMLKLTWTDRAAETAOL 725
Db 433 A-----GKQEADEVILGKAIR-----AMGPEAVLNLVPLNLAKPVKQGPGRW 476
Qy 726 GLVVDYDT---LLKLSVDSVTVSADGTRALVEATLESACLSDLVHPENNATDVRTYTR 781
Db 477 PLRDYTSNTNLAHFKSELVPLSALMFORVIE-----HQEKNIHIKIFETV 523
Qy 782 YEVPWS 787
Db 524 VQGIWS 529

RESULT 12
T30312
pilin biosynthetic protein - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 04-Mar-2000
C:Accession: T30312
R:Whitchurch, C.B.; Young, M.D.; Hobbs, M.; Mattick, J.S.
submitted to the EMBL Data Library, November 1996
A:Description: Pseudomonas aeruginosa chemotactic transduction genes pilA, pilB, pilC and
A:Reference number: Z20819
A:Accession: T30312
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-836 <WHI>
A:Cross-references: UNIPARC:UPI00001792CF; EMBL:U79580; NID:G3241967; PID:G3241968;
C:Genetics:
A:Gene: pilA
C:Superfamily: Pseudomonas aeruginosa pilin biosynthetic protein

Query Match 3.2%; Score 128.5; DB 2; Length 836;
Best Local Similarity 20.0%; Pred. No. 3.4;
Matches 158; Conservative 106; Mismatches 292; Indels 235; Gaps 35;

Qy 124 DALISRQILQAAACETLSNPRSRREYNEGLLDBDEATVITDVPWMDKVPFALCVLQEGGE- 182
Db 24 ETLKQARQALEAFVENPQDPTRMR-----FCLTYV--HQVQGTLMVVEFYGAA 69
Qy 183 --TEIVLVRVCEALLKERLPKSFQDQVVLVMALEFLVDSRDAMALDPPDF-----ITGE 234
Db 70 LLAERMEQLVQALLDGVRPNQGEALEVLMQAILQLPVYLDRIQTARRDLPMVPLFLNDL 129
Qy 235 FVEEALKLLOEGGASSLAPDL---RAQID-----ETLEETTPRVLELLGLP 278
Db 130 RAARGEKLLSE--TSLFAPDLSQRPQDGEAIAQLRTDELGLGLRLKRLQRTQOMALVGLL 187
Qy 279 LGDDYAAK-----RLNGLSGVRNI--LWSVGGGGASALVGGLTRKTFMNEAFLRMT 327
```


Db 188 RQDVATSLGYLARVYARLEGICREAPLGPLWSI-----ASGLVEGLANGSVNSASVR-T 242
Qy 328 AAEQVD-----LFVATPSNI-----PAESFEYVEVALALVAQAFIG 363
Db 243 LLRLDRELKRLVEQADGLNQAADELKVNLLFYVAKAPSQSPRILAL----- 291
Qy 364 KKPILLQDA--DKQFOLOQAQKVMAMEIPAM-----LYDTRNNWEIDFGLERG-- 409
Db 292 KEQYRLDEALPDHETVDAERARLAGDRDAMRSVVGALCEELVRIKDSLDLFVRSRGRHP 351
Qy 410 --LCALL--1GKVDCEKRWGLDSEDSQYRNPAIVEFVLENSNRDND-----DLPLGLCKL 461
Db 352 SELDALLAPLKQIADTAVLGVFGCPQPKVILDQDVIHALAQGRREFSDAILMDVAGALLY 411
Qy 462 LETWLAGVVFPP-RFRDTKDKKFKLGDYD-----DPMVLSYL----- 497
Db 412 VEATLAGMAGPGDERNSESRUPTTDAQIHLVKEARNGLQKQADAIIEFTASQWNHE 471
Qy 498 -----ERVEVQGS-----PLAAATMARIGABHVKASAMQALQKVPF----- 535
Db 472 HLARVPELLTQVRGGLAMIPLEAATLTLETENRYIQEQLL--ARKAVPDWQSLDTLADAI 529
Qy 536 ---SRVTDNSARPKDVQETVFSV--DPVGN-----NVGRDGEPCGVFTAEAVRPSNP 583
Db 530 TSVEYLLERLSEHASQSDILDVADSLANLGYTLKPNSSAPAEPL----- 577
Qy 584 ETNDVAIRAGVSSSYDETTVMESVADMLKEASVKILAAGVAIGLISLFSQKFLKSSSS 643
Db 578 -SGPAAIESPAAPERPEAVVE--VAETAQPPADTAPAEA----- 616
Qy 644 FORKDWSSMESDVATIGSVRADDSALPRMDARTAEINIVSKWKIKSLAFGPDHRIEML 703
Db 617 --REDAPQLASDNDWTLGCVAPDAGE--PSLDL-----ALDPLDSDSAEVP 658
Qy 704 PEVLGRMLKIWDRAEAETLAQLGVVYDYLTKLSVDSVTVSAD-----GTRALVEATLES 758
Db 659 PAVPECRE-----TPPOSTAPARSLDDFSLDEIDLGLDLPADAPASGPAALADWSLPE 714
Qy 759 SACL-SDLVHP 768
Db 715 QWGLGDDLAQ 725

RESULT 13
S70368
DNA polymerase I - Bacillus stearothermophilus
C;Species: Bacillus stearothermophilus
C;Date: 06-Dec-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C;Accession: S70368
R;Riggs, M.G.; Tudor, S.; Sivaram, M.; McDonough, S.H.
Biochim. Biophys. Acta 1307, 178-186, 1996
A;Title: Construction of single amino acid substitution mutants of cloned Bacillus : team
A;Reference number: S70368; MUID:96283813; PMID:8679703
A;Accession: S70368
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-876 <RIG>
A;Cross-references: UNIPROT:P52026; UNIPARC:UPI000016E912; EMBL:L42111; NID:9806280; PID
C;Genetics:
A;Start codon: TTG
C;Superfamily: DNA-directed DNA polymerase I

Query Match 3.2%; Score 128.5; DB 2; Length 876;
Best Local Similarity 21.0%; Pred. No. 3.6;
Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;

Qy 99 HFLTDGIRAFARVSKPPQFGFSDDALISRRQILQAACETLGNPRSRREY-NEGLLDDE 157
Db 107 HYEADDICTMAAAREE---GFAVKVISGDRDLTQ-----LASPQVTVITKKGIIDIE 158
Qy 158 EATVITDV-----PMDKVPF-----ALCVLOEGGETEIVLRVG 190
Db 159 SYTPETVVEKYGLTPQIVDLKGLMGKSDNIPGVPGIGEKTAVKLLKQFGTVENVLASI 218

Qy 191 EALLKRLPKSFQK--DVLVLMALAFDVSQDA---MALDPPDPFITGYEFVEEALKLQOE 245
Db 219 DEIKGKELKENLRQYRDAL-LSKLAAICRDAPELVLTLD--DIVYKGEDREKVVALLQOE 275
Qy 246 EGASSLAPDLRAQIDE-----TLBEIIPRYVLELLGL---PLGDDVYAAKRLNG- 290
Db 276 LGQSFLDKMAVQTDGEKPLAGMDFAIADSVTDEMLADKAALVVEVVDYNNHHIAPVGI 335
Qy 291 -----LSGVNRLWSVGGGASALVGGLTREKPMNEAFLKMTAAE----- 330
Db 336 ALANERGRFPLRPETALADPKFLAM-----LGDEYKKTMTFDSKRAVALKWKGI 385
Qy 331 -----QVDLFVATPSNIPAES-----FEYVEVALALVAQAFIGK-----KPHLLQ 370
Db 386 ELRGVVFDDLAAAYLLDPAQAAGVAAVKMHQVEAVRSDEAVYKGAKRKTVPDPTLAE 445
Qy 371 DADKQFOLOQAQKVMAMEIPAMLYDTRNNWE-IDFGLERGLCALLIG-----KVDECRM 423
Db 446 HLAR-----KAAAIWALEBPLMDLRRNEQDRLLTELEQPLAGILANMEFTGVKVDTKRL 500
Qy 424 -WGLD-SEDSQYRNPAIVEFVLENSNRDNDLPLGLCKLETWLAGVVFPRFRDTRDKK 481
Db 501 EQGAELTEQLQ---AVERRIYELAQEENINSP---KQLGT---VLFDKLQLPVLKK 549
Qy 482 FKLGDYDDPMVLSYL-ERVEVQGSPLAAAAATWARIAGBHVKASAMQALQKVPFSRYTD 540
Db 550 TKTG-YSTSADVLEKLAPHHEIVE--HILHYRLQGLQSTVIE-GLLKVVHVPTGVKVTM 605
Qy 541 RNSAEPKDVQET--VFSVDVPGNVGRDGEPCGVFTAEAVRPS-----NFEINDVAI 590
Db 606 FNQA-----LTQTRLSSVEPNLQNIPIRLEGRKIROAFVPSFPDWLIFAADYSQIELRV 661
Qy 591 RAGVSSSVDETTVE-----MSVADMLKE---ASVKILAAGVAIGLISLFSQ 634
Db 662 LAHIAE---DDNLEAPRRGLDHTKTAMDI FHVSEEDVTANMRQAKAVNFGVYGLSD 718
Qy 635 KYFLKSSSFQRKDWSSMESDVATIGSVRADDSALPRMDARTAEINIVSKWKIKSLAF 694
Db 719 -YGLAQNLTNITKEAAEFIERYPASFPQVK-----QYMDNIVQE-AKQKGYVT 764
Qy 695 GPDHRIEMLPEVLDRGM-LKIWDTRAA-ETAQLGLVYDYLTKLSVD-SVTVSADGTRA- 750
Db 765 TLLHRRYLPDITSRNFNVRSAERTAMNTPIQGSAAD-IKKAMIDLSVRLREERLOAR 823
Qy 751 -----LVEATLRESACLSDLVHPE 769
Db 824 LLLQVHDELILEAPKEIEIRLCRLV-PE 850

RESULT 14
E83594
still frameshift probable component of chemotactic signal transduction system PA0413 [im
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C;Accession: E83594
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bri
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: E83594
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2472 <STO>
A;Cross-references: UNIPROT:Q91696; UNIPARC:UPI00000C5045; GB:AE004478; GB:AE004091; NID:
C;Genetics:
A;Gene: PA0413

Query Match 3.1%; Score 127.5; DB 2; Length 2472;
Best Local Similarity 20.0%; Pred. No. 22;

| Matches | 159; | Conservative | 105; | Mismatches | 291; | Indels | 239; | Gaps | 36; |
|---------|------|---|------|------------|------|--------|------|------|-----|
| Qy | 124 | DALISRRQILQAACETLSNPRSRREYNELGLDDEEATVITDPWMDKVPFALCVLQEGGE- | 182 | | | | | | |
| Db | 19 | ETLKQKQALEAFVENPQDPTRMR | 64 | | | | | | |
| Qy | 183 | --TEIVLRVGEALLKERLPKSPKQDVILVMALAFLDVSRDAMALDPPDF-----ITGYE | 234 | | | | | | |
| Db | 65 | LLAEMEOVLQALLDGRVFNQGEALEVLMQAILQFVYLDRIQTARRDLPMVVLPLNDL | 124 | | | | | | |
| Qy | 235 | FVEEALKLLOEGEGASSLAPDL---RAQID-----ETLEEITPRVYLELGLP | 278 | | | | | | |
| Db | 125 | RAARGEKLLSE--TSLFAPDLISQRPQLDGEAIAQLRTDELGGLLRKLRTQOMALVGLL | 182 | | | | | | |
| Qy | 279 | LGDDYAAK-----RLNGLSGVNRI--LMSVGGGGASALVCGGLTREKPMNEAFIRMT | 327 | | | | | | |
| Db | 183 | RNQDVATSLGYLARVYARLEGLCREAPLGLWSI-----ASGLVPEGLANGSVVNSASVR-T | 237 | | | | | | |
| Qy | 328 | AAEQVD-----LFVATPSNI-----PAESFEVYEVALALVAQAFIG | 363 | | | | | | |
| Db | 238 | LIQLDRELKELVFGQDGLNQAPDELVKULLFYVAKAFSSQSPRIAL----- | 286 | | | | | | |
| Qy | 364 | KKPHLLQDA--DKQFOQLQAKVMAEIPAM-----LYDTRNNWEIDFGLERG-- | 409 | | | | | | |
| Db | 287 | KEQYRLDEALPDHETVDAERARLAGPDRDAMRSVVGALCEEELVRIKQSLDFVRSDRGHP | 346 | | | | | | |
| Qy | 410 | --LCALL--ICKVDECRMWGLDSEDSQYRNPATVEFVLENSNRDDND-----DLPLGLCKL | 461 | | | | | | |
| Db | 347 | SELQALLAPLQKQIADTTAVLGLGQGFQPKVILQDLQVIHALAQGRREPSDAILMDVAGALLY | 406 | | | | | | |
| Qy | 462 | LETWLAGVVFPP-RPRDTKDKKFKLGDYVD-----DPMVLSYL----- | 497 | | | | | | |
| Db | 407 | VEATLAGWAGPGDERNSEESLPTTQVIAQHOLVKEARNGLEQAKDAIETFTASQWNHE | 466 | | | | | | |
| Qy | 498 | -----ERVEVQGS-----PLAAATATWARGAEHVKASAMQALQKVPF----- | 535 | | | | | | |
| Db | 467 | HLARVPELLTQVRGGLAMIPLERAATLTETCNRYIQEQLL--ARKAVPDMQSLDTLDAI | 524 | | | | | | |
| Qy | 536 | --SRYTDRNSAEPKDVQETVFSV--DPVGN-----NVGRGCEPGVFTAEAVRSENF | 583 | | | | | | |
| Db | 525 | TSVEYVIERLSESHASQSDLLDVAEDSLANLGYTLKPNSSAPAEPLG----- | 572 | | | | | | |
| Qy | 584 | ETNDYAIRAGVSESSVDTETVMSVADMLKEASVKILAAGVAIGLISLFQKTYFLKSSSS | 643 | | | | | | |
| Db | 573 | -SGPAAIESPAABEPERPEAVVE--VAETASQPPADTAPAEAA----- | 611 | | | | | | |
| Qy | 644 | FORKDMVSMESDVATTIGSVRADDSSEALPRMDARTAEINIVSKWQIKSLAFGPDHRIEM- | 702 | | | | | | |
| Db | 612 | --REDAPQLASDNNTLGEVAPDAGE--PSLIDL-----ALDPLDSDAEVP | 653 | | | | | | |
| Qy | 703 | --LPEVLGDGMKLTWDRAAETAQLGLVYDYVITLLKLSVDSVTVVSAD-----GTRALVEAT | 755 | | | | | | |
| Db | 654 | PALPEVVEESGQQSITAPARSILD-----DPSLDETDLSGLDLPADAA PASGPAALADWS | 708 | | | | | | |
| Qy | 756 | LEESACL-SDLVHP | 768 | | | | | | |
| Db | 709 | LPEQWGLGDDLAOP | 722 | | | | | | |

RESULT 15
T30853
antigenic heat-stable protein - Rickettsia japonica
C/Species: Rickettsia japonica
C/Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 09-Jul-2004
C/Accession: T30853
R/Uchiyama, T.
submitted to the EMBL Data Library, May 1997
A/Description: Sequencing of the gene encoding the antigenic heat-stable 120-kilodalton protein from *Rickettsia japonica*.
A/Reference number: Z20905
A/Accession: T30853
A/Status: preliminary; translated from GE/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1018 <UCH>

A,Cross-references: UNIPROT:Q9AJ79; UNIPARC:UPI0000135607; EMBL:AB003696; NID:d11066377;

Query Match 3.1%; Score 126.5; DB 2; Length 1018;
Best Local Similarity 19.2%; Pred. No. 6.2;
Matches 128; Conservative 109; Mismatches 265; Indels 163; Gaps 28;

| | | | |
|----|-----|--|-----|
| QY | 115 | KPQPGFSDALISRRQILQAACETLSNPRSRREYNEGLLDDEERATVITVPWDKVPFGAL | 174 |
| Db | 374 | EPQYKQVPPITPTNQPLQF--ETSQMPQS-QQVNPLLN--AATALSGSQDILLNYYN | 428 |
| QY | 175 | CVLOEGGETEIVLRUGEALIKERLPKSFQDQVLMVA--LAPLDYSRDAMALPPDPFITG | 232 |
| Db | 429 | AGLTG--EIDLKEARATAILNDK-----KSDIAEKQANIILAENTVNNKNTTPDAKVAG | 481 |
| QY | 233 | YEFVEALKLLOBEGASSLAPDLR-----AQIDETLEEITPRYVLEL---LGLP | 278 |
| Db | 482 | VNAVLETIKNDQN-----TPDLEKSKMLEATVAIALNSENLEPKQKQKQILKAVDVGLS | 535 |
| QY | 279 | LGDDYAAKELNGLSGVNRNIIILSVGGGASALVGLTREKFWNEAFRL--MTAAEQVDLFLV | 336 |
| Db | 536 | LKDD-----ASRAAIDGIT-----DAVKSNLSTEDKGTMTFI | 568 |
| QY | 337 | ATPSNIPAESFEYVEVALVAQAFIGKPKHLLQDADKQFOLOQAKYMA-----ME | 388 |
| Db | 569 | AVGDKVNVSELSNAEKQLGLSVLKGVAEQVLSPAQQQLMQQNLDKITAEQTKKDTIKK | 628 |
| QY | 389 | IPAMLYDTRNWE-----IDFGLERGLCALLIGKVDECRMWLGSDSDSQYR | 435 |
| Db | 629 | VNDILFDPLSNLTSLKNTQIALLISNVLDPATAEVKGEIIQEITNTVAGSSLEAHD---- | 684 |
| QY | 436 | NPAIVEFVLN--SNRDNDDLPGLCKLLETWLAGVVPFRPDTDKDKFKGLGDDYDDPMV | 493 |
| Db | 685 | KAAIINGISSETIATHSDTSLSPNKALIMASAEKGIA-----ESQNLDP--RELMT | 734 |
| QY | 494 | LSYLERVEVQGSPLAAAAATMARIGAHHVKASAMQALQK---VFPSTRYDTRNSA----- | 544 |
| Db | 735 | KGLVDGIVEGKGPEITTKAVSSGIDNSINIDSEKEALKKAKDAASEAALDRETQNLTEGL | 794 |
| QY | 545 | -----EPKD-----VOETVESDPVGNVGRDGEQVFAE----- | 575 |
| Db | 795 | KQNTIEHKPHDDIYINKREVINAIVPVI BALEKSKPEFVSAERIVQETSGILNANSIKL | 854 |
| QY | 576 | AVRPSNFETNDYAIRAGVSESSVDETTVMSSVADMLKEASVKILAAQVAGTILSIFSOK | 635 |
| Db | 855 | AVEKYNPF-----RAMLSPNGNLKLEK-----KEESI KKVDE-----LVKAFGTG | 896 |
| QY | 636 | YFLKSSSSFPQKDMVS-----SMESDVATIGSVRADD---SEALPRMDARTAE-NIVSKWOK | 688 |
| Db | 897 | SSTEQSQFIKANLDDDKTSLKEIRLQTIKDLQEKRAEAIENPSVKTEDVRVVSQSK | 956 |
| QY | 689 | IKSLA | 693 |
| Db | 957 | LKPIS | 961 |

Search completed: December 4, 2005, 12:50:55
Job time : 49 secs

Search completed: December 4, 2005, 12:50:55
Job time : 49 secs